1965

On unbiased estimation in variance component models

Rodney Peter Basson
Iowa State University

This dissertation has been microfilmed exactly as received 65-12,461

BASSON, Rodney Peter, 1936-
ON UNBIASED ESTIMATION IN VARIANCE COMPONENT MODELS.

Iowa State University of Science and Technology, Ph.D., 1965
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan

ON UNBIASED ESTIMATION IN VARIANCE COMPONENT MODELS

by

Rodney Peter Basson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1965

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

I. INTRODUCTION

By a variance component model we shall mean a statistical model with vector/matrix representation of the form

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=r+1}^{k+1} X_i \beta_i$$

where y is a n x 1 vector of observations, X_i's are fixed, known, n x p_i matrices whose elements may be either classification or functional parameter coefficients, γ_i's are fixed parametric vectors and β_i's are random vectors satisfying $E(\beta_i \beta'_j) = 0(i \neq j)$, and $E(\beta_i \beta'_i) = V_i$. Sometimes, but not always, the assumption that $E(\beta_i \beta_i') = I \sigma_i^2$ will be made as will the assumption of normality of β_i effects. By appropriate restrictions we can make our model conform to the classification of Eisenhart (1947) or to some of the cases mentioned in the extended classification of Tukey (1949). Thus Model I (of Eisenhart (1947)) refers to the case where k=0, all X_i's are 0 or 1 and there is only one random vector in our model. In this case where X_i's (i=0,...,r) refer entirely to classification parameters the model represents an analysis of variance (A.o.V.) model, while if X_i's consist of both classification and functional parameters this special case represents an analysis of covariance (A.o.C.) model. Model II (of Eisenhart (1947)) refers to the situation where r=0, $\gamma_0=\mu$, $X_0=j_n$ i.e., an n x 1 column vector of ones, and the β_i's are assumed to be independently normally distributed with covariance matrix I σ^2, assuming I of appropriate dimension for each i. Model III of Tukey (1949) is the pure finite sampling model and is similar to Model I except that the populations from which the random errors are drawn are assumed to be finite, so that correlations are induced into the covariance structure of y.

The minimum variance (M.V.) unbiased point estimation problem usually seems to be viewed in two separate parts that are seldom discussed together. These are: a) The estimation of fixed effect parameters in models with correlated or uncorrelated errors at which time nothing much is said about estimates of components of variance except in the special case of general linear hypothesis. b) The estimation of variance components in random effects models at which time nothing is said about the estimation of fixed parameters in the model. One of our aims was to give consideration to these problems simultaneously, but in point of fact, a reasonable solution is found in only a small number of cases. We consider a "mixed" model and it appears to be the case that only in certain specially "balanced" cases at opposite ends of the spectrum of the general model that we are considering are estimates with "optimum" properties available. Naturally the optimum properties depend somewhat on the assumptions made at the start, and "balance" is a term which needs clarification.

In experimental design situations when all factors are not in fact random, but some are fixed, statisticians in an attempt to obtain an approximate solution for estimators of variance components have suggested answers which are motivated by either the completely random "balanced" or the general linear hypothesis case and then modified in some fashion. We have in mind, for example, the well-known rules for writing down expected mean squares for the two-way mixed model with interaction, given, amongst others, by Bennett and Franklin (1954); the motivation for some modifications has not always been as clear as it could have been.

Some of the contributions made by the present work are:

a. An attempt to unify the knowledge on M.V. unbiased point estimation in a variety of models and to extend the available techniques for obtaining best, i.e., M.V., estimators of components of variance to some designs, including both random and mixed model ones for both the infinite and the finite model where currently either little or nothing is known about the properties of least squares estimators.

b. The development of a classification for variance component models (of the type defined above), that enables us to tell by a simple computation whether estimators with good properties are available for any design in question. Although investigated for an infinite model, we show grounds for believing that this classification will be useful in determining whether the best quadratic unbiased (b.q.u.) estimator property can be inferred for finite model estimators.

c. In most of the cases that can be represented by the variance component model above, it is the case that U.M.V. unbiased estimators do not exist. We have obtained some new results on the variances of quadratic forms arising in either mixed or random models and which are not dependent on the normality assumption. Naturally however, they reduce to simpler forms in that case. We indicate briefly how these results are necessary to solve the question of which estimator we should choose among reasonable alternatives in virtually any cases that fall outside the rather small "balanced" class in which U.M.V. estimators do exist.

II. REVIEW OF LITERATURE AND INITIAL RESULTS

Due to the broad nature of the topic and the fairly extensive literature thereon, this review is rather long. It may be divided quite naturally into two sections:

A. Estimation of components of variance

B. Minimum variance estimation of regression parameters in models with correlated errors.

A. Estimation of Components of Variance

The variance component estimation problem in a model of the type

$$y = j_n\mu + \sum_{i=1}^{k+1} X_i \beta_i \quad \text{(II.A.1)}$$

is theoretically regarded as solved if we are satisfied to make the assumptions of no correlation of the different β_i type vectors, i.e. $E(\beta_i \beta'_j) = 0 (i \neq j)$, normality of distribution of β_i type vectors and $E(\beta_i \beta'_i) = I\sigma_i^2$. Actually, however, maximum-likelihood provides an answer without resorting to iterative techniques only in "balanced" models even when all assumptions are made. In practice then, if we are to obtain estimators with reasonably powerful small sample properties, in contrast to ones which are only asymptotic, the assumption of some sort of balance is a very definite requirement.

It seems everybody has his own definition of balance. For example Crump (1951) says "A multiple classification is balanced if all of the classes or subclasses of any chosen rank contain the same number of observations." Tukey (1956) has another view which relates to the expectations of different lines of the A.o.V. table and their behavior when arbitrary changes are made in the contributions in any particular line. We find it convenient to point out at this early stage that we shall distinguish two types of balance, namely

$$\text{balance}_1$$

and

$$\text{balance}_2 .$$

In general linear hypothesis models (frequently written $y = X\gamma + e$) the notion of balance$_1$ from one point of view requires the existence of an easily found solution to the normal equations $(X'X)\gamma = X'y$. Apart from the desirable feature of simplifying the analysis this notion of balance usually ensures a structured covariance matrix and often equal information on estimates of treatment differences. This appears to be the main advantage of a balanced$_1$ design. From the viewpoint of analysis, of course, because of the Gauss-Markoff theorem, there is no need to restrict consideration to balanced$_1$ situations in order to obtain estimators with the best linear unbiased (b.l.u.) property. There is not and never was any use for the concept of balance$_1$ in variance component models, but this is not generally recognized or agreed upon. We shall refer to the model (II.A.1) where only β_i's are random, $E(\beta_i \beta_j') = 0$ $(i \neq j)$ and $E(\beta_i \beta_i') = I\sigma_i^2$ as Model II. A model representation (II.A.1) will be said to be balanced$_2$ if and only if $X_i X_i' X_j X_j' = X_j X_j' X_i X_i'$ $(i, j=0,\ldots k+1)$. By Theorem 3 of Chapter III this means that there exists a known orthogonal matrix which diagonalizes the covariance matrix

$$V = X_1 X'_1 \sigma_1^2 + \ldots + X_{k+1} X'_{k+1} \sigma^2_{k+1} \ ,$$

regardless of the actual values of the unknown parameters. In fact, the orthogonal matrix must simultaneously diagonalize

$$X_i X_i' (i=1,\ldots,k+1) \ .$$

We shall see that the balance$_2$ concept is a natural one in variance component models. There is no obvious reason why any balanced$_1$ design should also be a balanced$_2$ design. In classification designs this may however be the case. It should be noted that balance$_2$ is not the same as the balance of Crump (1951) which in turn is not balance$_1$, yet all three classes do to some extent overlap. Balance$_2$ is not the same as the balance of Tukey (1956) either. To be more specific, Tukey (1956) deals with a Model III situation and regards the b.i.b. with treatments and blocks random as balanced.

However, if we represent the random case of the balanced incomplete block design (b.i.b.d. in the sequel) in the form (II.A.1) with treatments and blocks random, then it is not balanced$_2$.

Crump (1947, 1951) was of the opinion that the estimates of various components obtained by regarding all parameters as fixed and equating observed and expected mean squares under the random model (and replacing negative estimates by zero) i.e. what are known as A.o.V. estimators, are maximum likelihood (M.L.) estimates when Model II is assumed and the classification is balanced, i.e., if all of the classes and subclasses of any chosen rank contain the same number of observations. Ignoring for now the problem of equating negative estimates to zero, this conclusion still needs qualification. For example it is not clear that the M.L. estimators equal the A.o.V. estimators and that either have any special merit in the design

$$y_{ij}= u+a_i+b_j+e_{ij} \qquad \text{(II.A.2)}$$

$$(i=1,2),\ (j=1,2);\ (i=3,4),\ (j=3,4)\ .$$

In fact due to lack of completeness we cannot infer M.V. for M.L. or any other estimators in this case. The situation in a b.i.b. is similar, though more complex. We shall go into this case in more detail later. The balance criterion of Crump (1951) appears to be of no help in sorting out difficulties such as the above. The introduction of the balance$_2$

definition seems to be a step in the right direction. Major credit for the concept, if not for the name, is due to Graybill and Hultquist (1961) who appear to have been the first to use the condition of balance$_2$ in a class of variance component models. They did not, however, point out that if this rather exacting condition does not hold and even sometimes when it does (as in II.A.2 above) we are in difficulty insofar as M.V. estimation is concerned. Daniels (1939) and Satterthwaite (1946) suggested estimators of the sampling variance of any estimated variance component in a balanced multiple classification under Model II. The sampling distribution of such estimates under Model II have been given by Pearson (1933) and discussed by Satterthwaite (1946) and Bhattacharyya (1945). Bross (1950) discussed and illustrated several approximate methods of obtaining confidence limits for estimated variance components. Wald (1940, 1941, 1947) has given a theoretical method (apparently not yet used in practice) by which exact confidence limits for any ratio of a variance component to the error component may be obtained. Wald (loc. cit.) used a general linear hypothesis approach to obtain mean squares. This method, of ignoring momentarily the real assumptions and using least squares, was spelled out and popularized by Henderson (1953). If we substitute an estimate of error for σ^2, Wald's (loc. cit.) method would also yield approximate confidence limits. Crump's (1951) pioneer article reported on some of the work that was getting under way, which regarded Model II as much too restrictive and which desired to obtain estimates under less restrictive conditions. Tukey (1950, 1956 , 1957), Hooke (1956a, and b) and Wilk and Kempthorne (1955, 1956) provide a small indication of work which included amongst its general aims the estimation of variance components (and such redefinitions of this concept as were found necessary) when the assumption of normality and drawing effects from infinite populations were replaced by drawing from finite ones.

It appears to be fair to say that a good measure of success was achieved in obtaining unbiased estimators for the balanced$_1$ structures, i.e., essentially those cases in which equal or proportional numbers were observed in the cells of crossed or nested structures. White (1963) has extended the rules for finding expected mean squares to some of the designed unbalanced structures. The similarities between the polykays (of Tukey) and the cap sigmas (of Wilk and Kempthorne) were pointed out by Zyskind (1958) and established by Dayhoff (1964b). The A.o.C., i.e. a non-classification model, has not been adequately resolved under a randomization model and Cox (1956) still appears to be the only one who has attempted to deal with this case. The approach in this randomization model is general enough to deal with a mixed model. The only claim made for the estimates obtained is that generally they are unbiased. Obtaining the variances of the variance components under a finite model in even the simplest balanced n-way classifications is a formidable problem which has been attempted by Dayhoff (1964a).

When we relax only the assumption of normality of distribution of the effects and instead regard them to be independent(0, σ^2_i) respectively, estimators with a best quadratic unbiased property are available in the following cases:

a. If in (II.A.1) $\beta_1, \beta_2, \ldots, \beta_k$ are all fixed unknown parameters. This is the well-known general linear hypothesis situation and constitutes one of the oldest parts of statistical theory. Least squares provides b.l.u. estimators of the parameters $\beta_1, \beta_2, \ldots \beta_k$ and the S.S. about residuals with a further condition on the fourth moment provides a b.q.u. estimator of the variance $\sigma^2_{k+1} = \sigma^2$. This result is due to Hsu (1938). We emphasize that the elements of the X_i's are classification type elements or functional parameter coefficients.

b. Graybill (1954) showed from first principles that the Model I mean squares provided b.q.u. estimators for functions of the variance components in the balanced nested classifications.

c. By making use of a different approach that was first suggested by Graybill and Wortham (1956), i.e. complete sufficient statistics if normality holds, Graybill and Hultquist

(1961) showed that within a certain class of situations described by the model (II.A.1) if finite fourth moments of β_i's exist and third and fourth moments for a given β_i are equal then the same estimators that are M.V. under normality are b.q.u. under the assumptions mentioned.

Best quadratic unbiasedness has not been claimed for most other estimators that are commonly used in variance component models. Notably this is the case for most unbalanced, i.e. not balanced$_2$ situations that arise in experimental design and where for some time now least squares has been used as an "approximate" method. Thus we have the situation that the same method is suggested for balanced$_2$ situations as well as for unbalanced ones, while optimum properties are available for least squares estimates only in the former case. Estimates would be "best" in both cases if the fitting constants method was an "optimal" method of estimation in variance component models. Our viewpoint is that there is in general no uniformly best method of estimation for the model (II.A.1) when amongst estimators that are based on all the information there is ambiguity about the set of estimators that should be used. In such a case one method of comparing a candidate with other available methods is by way of the variances of alternative estimators. Empirical studies by Bush and Anderson (1963) indicate that the fitting constants method or Henderson's (1953) method 3 is a "best on average" method over a fairly wide range of conditions of unbalance in a two-way classification. In a later chapter we shall advance some further argument that appears to single out this method as the best available in b.i.b. designs.

When we are prepared to make the assumptions of normality and independence we find that for some balanced$_2$ situations, the model 1 mean squares can be shown to consist of a complete sufficient set of statistics for the mean and the variance components. Graybill and Wortham (1956) first pointed out that this enabled one to call upon the well-known Rao-Blackwell theorem, (Rao (1945), Blackwell (1947)) dealing with complete sufficient statistics to establish properties for unbiased estimators. Graybill and Hultquist (1961) prove a number of theorems that have helped in the understanding of the classification model cases when normality is assumed and the model is balanced$_2$. Assuming normality, Weeks and Graybill (1961, 1962) Kapadia (1962) and Kapadia and Weeks (1963) have obtained what they call minimal sufficient statistics for b.i.b.d.'s and p.b.i.b.d.'s with random blocks and treatments. At time of writing we have minimal sufficient statistics for virtually all incomplete block designs but no method of using these to obtain best estimates. We shall give an example showing how closely the method 3 estimates of Henderson (1953) are related to the set of minimal sufficient statistics from which it is presumably the view of some that

M.V. estimators may one day be constructed. We have our doubts about this, and present our views.

There has been one fairly detailed investigation of the testing for significance of variance components, namely that of Herbach (1957). As is usual in such cases, normality of effects was assumed. Herbach (loc. cit.) showed for example that all F-tests used in testing hypotheses in a two-way classification determine uniformly most powerful (U.M.P.) similar tests although, unlike in the case for Model I, they are not likelihood ratio (L.R.) tests. In the one-way balanced$_{1 \text{ or } 2}$ classification however for all practical purposes the test is an L.R. test and is U.M.P. as well. Herbach's (1957) method was unable to demonstrate that the standard F test for $\sigma^2_{AB}=0$ in the two-way balanced$_2$ case, is a U.M.P. similar test, and this deficiency was remedied by an extension of the completeness lemma supplied by Gautschi (1959) who had worked independently on the problem. Gautschi (loc. cit.) proved the following Lemma 1:

$$\text{Let } \mathcal{B}^t = \left\{ P^t_\theta ; \theta \in D \right\}, \quad t = (t_2, \ldots, t_r), \ \theta = (\theta_2, \ldots \theta_r)$$

$$\mathcal{B}^{t1} = \left\{ P^{t1}_{\theta_1, \theta} ; (\theta_1, \theta) \in \mathcal{D}_1 \times \mathcal{D} \right\} \quad \theta_1 \text{ real be two}$$

families of probability measures on Borel sets of the Euclidean space E_{r-1} and the real line E_1 respectively, having the densities

$$p_\theta(t) = c(\theta)h(t_2,\ldots,t_r)e^{\theta_2 t_2+\ldots+\theta_r t_r}$$

$$p_{\theta_1,\theta}(t_1) = c(\theta_1,\theta)e^{g(\theta)t_1^2+\theta_1 t_1}$$

with respect to Lebesque measure. If $\mathcal{D}_1$ is the real line and $\mathcal{D}$ a Borel set in E_{r-1} containing a non-degenerate (r-1)-dimensional interval then the family of product measures $\mathcal{B} = \left\{ P^{t_1}_{\theta_1,\theta} \times P^{t}_{\theta}; (\theta_1,\theta)\ \mathcal{D}_1 \times \mathcal{D} \right\}$ is strongly complete (in the sense of Lehmann and Scheffe (1950). Imhof (1960) extends this lemma to the multivariate case. The state of knowledge at this time appears to be that in an n-way balanced situation all tests of (n-1) and (n-2) factor interactions are U.M.P. similar tests.

Workers have not ignored the unbalanced case in recent years. For details the reader may refer to Hammersley (1949), Henderson (1953), Tukey (1957), Searle (1956, 1958, 1961) Gates and Shiue (1962), Gower (1962), Robertson (1962), and Bush and Anderson (1963) amongst others. The above selection illustrates the complexity of the algebra we sometimes involve ourselves in when trying to obtain variances of estimators and the sometimes surprisingly elegant rewards to be reaped from hopeless seeming cases.

Crump's (1951) paper in retrospect suggests the need for trying and comparing different methods of estimation. It is interesting to note that in the unbalanced (i.e., not balanced$_2$) designs the normality assumption if made has not to date been used to form a maximum likelihood solution. Computational difficulties are cited as the reason. Considerable work, some of which is empirical, has been done comparing some other different methods that are available. A whole new aspect has in fact been opened up by this work which is concerned with determining the best form of unbalanced design to estimate particular variance components when it can be assumed that something is known about the relative sizes of the variance components. The reader is referred to Bush and Anderson (1963) for further references on this aspect. Ignoring for now the moot question of motivation, we interpret these findings as follows: given a balanced$_2$ design, there is no ambiguity about the best statistical analysis. In an unbalanced design there is no uniformly best method but Henderson's (1953) method 3 (the fitting constants method) is among the best "on average" in the 2-way design without interaction. However, if anything at all is known about the relative sizes of the variance components, the balanced$_2$ design may well be less efficient for estimating the random components than some unbalanced design.

In general prior knowledge is not available; we concern ourselves primarily with that case, and try to extend the conclusion beyond the two-way design, and to mixed models, a venture in which we meet with only partial success. Until now we have been primarily concerned with the completely random model. One might perhaps have expected to find a comparable amount of work on the mixed model. There has been a good deal of work, but it has not concerned itself with minimum variance estimation in the mixed model. Thus for example the work of Wilk and Kempthorne (1955, 1956) is general enough to include the case of the mixed model. However, they start with the usual Model I mean squares and make no claim of minimum variance for the estimates they obtain. Scheffè (1956, 1959) likewise makes no claim of M.V. for such estimators as he derives. M.V. estimators in a mixed model are obtained by Imhof (1960) who derives a set of sufficient statistics in a three-way classification with one factor fixed and two random and proves M.V. for the Bennett and Franklin (1954) type mean squares in that case. The Scheffè model postulates non-zero correlations between different random factors and is therefore outside the framework that we consider.

In the unbalanced case we note that the fitting constants method does apply to a mixed model. However, no formulas have been given for obtaining the variances of the variance components. David and Johnston (1951, 1952) make use of symmetric functions to obtain formulae for the cumulants of quadratic forms in fixed, mixed, and random models. We extend these formulae to apply to estimators given by the fitting constants method in a mixed model, and choose to find the variances by making use of the concept of a conditional inverse of a matri . Bush and Anderson (1963) seem to have been the first to solve the variance problem (for the completely random model) by another method that assumes normality. Although the results of David and Johnston (1951, 1952) are applicable they have not been put to use. To conclude section A of this review we mention some work by Thompson (1962) and Thompson and Moore (1963) on the estimation of non-negative estimates of variance components that has some intuitive appeal. It is unfortunately the case that the "pool the minimum violator" algorithm that they describe and which gives a straightforward procedure for solving for variance components is only applicable to the designs which have "rooted tree" form. In brief this means that the three-way and higher-way classifications are not included in the theory that they describe. This severely restricts the field of application of the method. The question of minimum variance or otherwise of such estimators is an open question; we give this method no consideration.

B. Minimum Variance (M.V.) Estimation of Regression Parameters

In the course of the above discussion we mentioned that estimation by least squares of the regression parameters in a general linear hypothesis model gave best (M.V.) linear unbiased estimators. Our concern here is with estimation in more general models which to date seem to have had attention only in the context of autocorrelated regressions. In this section we review some of the work in this area.

The work described does not utilize structure in the errors, so that it is convenient to represent the partitioned model (II.A.1) by

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=y+1}^{r+1} X_i \beta_i$$

or by

$$y = X\gamma + \omega \,, \qquad \text{(II.B.1)}$$

where γ denotes the fixed set of parameters while ω is a column vector with covariance matrix V.

The problem of estimation of regression parameters when the errors are correlated was first discussed by Aitken (1934); there has been a good deal of discussion since then that we do not propose to cover here. We have decided to confine attention largely to the papers by Watson (1955), Magness and McGuire (1962) and Golub (1963). In general these authors are concerned with the determination of bounds to the ratios of variances of minimum variance (M.V.) and least squares (L.S.) or weighted least squares (W.L.S.) estimates and showing under what conditions some or all of these methods would give estimates having variances of similar magnitude. In the cases they discuss, as in design situations, M.V. estimation is generally not possible. If L.S. estimation is almost as good, this would be nice to know.

The following theorem is proved by Magness and McGuire (1962). The W.L.S. and M.V. estimates of γ in

$$y = X\gamma + \omega$$

where $X'X=I$, have identical covariance matrices if and only if the subspace spanned by the p columns of X coincides with the space spanned by p of the eigenvectors of ρ (rho) the correlation matrix of y, in which case both covariance matrices are similar to diagonal matrices whose elements are the corresponding eigenvalues of ρ (rho). The result that L.S. estimators of γ are M.V. when column vectors of X are eigenvectors of the correlation matrix is mentioned also by Watson (1955). Magness and McGuire's (1962) work was, however, done without awareness of Watson's (1955) results. Zyskind (1962) and Zyskind et al. (1964) have given a more general formulation that does not require the restriction $X'X=I$. It follows from these results that the minimum variance estimate of μ in a random balanced$_2$ partitioned model of the form represented by

$$y = j_n\mu + \sum_{i=1}^{k+1} X_i \beta_i \qquad \text{(II.B.2)}$$

agrees with the W.L.S. estimate in which all weights are unity, or the estimate given by least squares ignoring the correlations. Alternatively, and more directly, we have

Proof: In a linear model $y = X\gamma + \omega$

$$\hat{\gamma}_{MV} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

and

$$\hat{\gamma}_{WLS} = (X'WX)^{-1}X'Wy \quad .$$

Then $\sigma^2_{\hat{\gamma}_{MV}} = (X'V^{-1}X)^{-1}$

and

$$\sigma^2_{\hat{\gamma}_{WLS}} = (X'WX)^{-1}X'WVWX(X'WX)^{-1} \quad .$$

In the model (II.B.2), $\gamma = \mu$, $X = j_n$, i.e. a n x 1 column of ones, and $XX' = J$, a n x n matrix of ones. Now the condition of balance$_2$, ensures that

$$VJ = \Sigma X_iX_i{}' \sigma_i{}^2 XX'$$

$$= XX' \Sigma X_iX_i{}' \sigma_i{}^2 = JV \quad .$$

If furthermore $W = I$, so that $\hat{\gamma}_{WLS} = \hat{\gamma}_{LS}$ then

$$\sigma^2_{\hat{\mu}LS}/\sigma^2_{\hat{\mu}MV} = X'V\ X\ X'V^{-1}X/(X'X)^2 = 1 \quad .$$

In general the vector space of X is not coincident with the vector space of the eigenvectors of V so that in general we cannot simplify det $\Sigma_{\hat{\gamma}LS}$/det $\Sigma_{\hat{\gamma}MV}$ (where det means "determinant" and this ratio is used to replace $\sigma^2_{\hat{\gamma}LS}/\sigma^2_{\hat{\gamma}MV}$ when constant parameters exceed one in number) to give 1 as in the above special case.

However, there is a fairly large well used class of situations in which simplification is possible. These are those mixed factor situations, which, if all factors had been random, would have been balanced$_2$. We demonstrate this for the mixed model (II.B.1) which can be represented by

$$y = X\gamma + \sum_{i=r+1}^{k+1} X_i \beta_i \qquad \text{(II.B.3)}$$

where $X = (X_0 X_1 \ldots X_r)$ is the coefficient matrix of all fixed factors, and commutativity of $XX' = \sum_{i=0}^{r} X_i X'_i$ and V (the covariance matrix in this model) holds.

Assuming that we reparametrize $X\gamma$ appropriately, to avoid singularity of $X'X$, to $\overline{X}\delta$ (say) where $\overline{X}$ is a matrix containing only some of the X_i's $(i=0,\ldots,r)$ we have (with $W = I$)

$$\det(\Sigma_{\hat{\delta}_{LS}})/\det(\Sigma_{\hat{\delta}_{MV}}) = \det(\overline{X}'V\overline{X}\overline{X}'V^{-1}X)/\det(\overline{X}'\overline{X})^2$$
$$= 1 \quad .$$

In particular then in a two-way with equal numbers and with blocks random, the $(y_{.j})$ are M.V. estimators of the parameters $u + \alpha_j$, and $y_{..}$ is a M.V. estimate of u. In general, for unbalanced situations, we do not have commutativity of $\overline{XX'}$ or XX' and V and the ratio

$$\det(\Sigma_{\hat{\delta}_{LS}})/\det(\Sigma_{\hat{\delta}_{MV}}) \qquad \text{(II.B.4)}$$

does not simplify to one. Watson (1955) and Golub (1963) both attempt to provide bounds for the ratio (II.B.4). After Golub (1963) we define $M_k = X'A^kX$, where A is a real positive definite matrix, and $u_k(X) = \det(M_k)$, where X has rank p, and the latent roots of A, ordered decreasingly, are $\lambda_1,\ldots,\lambda_n$; let the condition number be $K_p = \lambda_1,\ldots,\lambda_p/\lambda_{n-p+1},\cdots,\lambda_n$.

Schopf (1960), proves the inequality following:

$$1 \leq u_{k+1}(X) \quad u_{k-1}(X)/u_k^2(X) \leq [(K_p^{\frac{1}{2}}+K_p^{-\frac{1}{2}})/2]^2 .$$

Using this Golub (1963) proves (when k=0) the following theorem. Let $y = X\gamma + \omega$ where X is an n x p matrix of rank p, γ is a vector with p components to be estimated, and ω is a random vector of n components with $E(\omega)=0$ and covariance matrix V. Let the weighting matrix be $W=FF'$, let λ_i's be the eigenvalues of $F'VF$ ordered decreasingly and put $\Psi = F'X$. Then since

$$\det(\Sigma\hat{\gamma}_{WLS})/\det(\Sigma\hat{\gamma}_{MV}) = \det(X'WX)^{-1}X'WVWX(X'WX)^{-1}/\det(X'V^{-1}X)^{-1}$$

$$= \det(\Psi'F'VF\Psi)\det(\Psi'(F'VF)^{-1}\Psi)/(\det\Psi'\Psi)^2,$$

it follows that

$$1 \leq \det(\Sigma\hat{\gamma}_{WLS})/\det(\Sigma\hat{\gamma}_{MV}) \leq [K_p^{\frac{1}{2}} + K_p^{-\frac{1}{2}}/2]^2 .$$

Watson (1955) obtained a similar bound. As far as our problem is concerned however there appear to be distinct limitations to this type of bound. The following example illustrates the difficulty.

Consider the model for a one-way random classification represented by

$$y_{ij} = \mu + a_i + e_{ij}$$
$$(i = 1,\ldots,a; \quad j = 1,\ldots,b)$$

where μ is the mean effect, a_i is a random effect assumed to be distributed about zero with constant variance σ^2_A, and e_{ij} is a random error effect assumed to be distributed about zero with constant variance σ^2_E and a_i and e_{ij} are assumed to be uncorrelated. The covariance matrix of the observation vector y has the form

$$\begin{bmatrix} \sigma^2_E I_b + \sigma^2_A J^b_b & & \\ & \sigma^2_E I_b + \sigma^2_A J^b_b & \\ & & \ddots \; \sigma^2_E I_b + \sigma^2_A J^b_b \end{bmatrix}$$

where I_b is the b x b unit matrix, J^b_b is a b x b matrix of ones and there are a diagonal blocks.

The largest latent root is $\sigma^2_E + b\sigma^2_A$ with multiplicity a while the smallest latent root is σ^2_E with multiplicity a(b-1). We rank the latent roots from largest, designated λ_1, to smallest, designated λ_{ab}. Since there is only one parameter (μ) in the model we define

$$K_1 = \lambda_1/\lambda_{ab} = (1 + b\, \sigma^2_A/\sigma^2_E) \; .$$

According to the theorem quoted above

$$1 \le \det \Sigma_{\hat{\delta}LS}/\det \Sigma_{\hat{\delta}MV} = \sigma^2_{\hat{\mu}LS}/\sigma^2_{\hat{\mu}MV} \le ((K_1^{1/2} + K_1^{-1/2})/2)^2$$

(where we have assumed the weighting matrix W = I for lack of knowledge of better weights). Conceivably then if we assume $\sigma^2_A \gg \sigma^2_E$, the upper bound may in fact be arbitrarily large, and knowledge of the bound would allow very little to be inferred about how well the variance of M.V. estimates are approximated by the variance of L.S. estimates, when in fact from other points of view e.g., Magness and McGuire (1962), or since the situation is balanced, $\sigma^2_{\hat{\mu}MV} = \sigma^2_{\hat{\mu}LS} = \sigma^2_E + b\;\sigma^2_A/ab$, and it follows that the variance of the L.S. estimator is in agreement with the variance of the M.V. estimator. When for example we consider the model

$$y_{ij} = \mu + a_i + \beta x_{ij} + e_{ij}$$

where β is an unknown fixed value and x_{ij} is a known fixed value, there are two parameters in the model and so

$$K_2 = \lambda_1 \lambda_2 / \lambda_{ab} \lambda_{ab-1} = (1 + b\, \sigma^2_A / \sigma^2_E)^2 \quad .$$

Use of the bound in this case to infer something about L.S. estimators seems to hold little promise. A further increase in the number of concomitants causes the bound to tell us even less. These considerations form the basis for the criticism that when other methods are uninformative and we need the bound most, it is liable to be uninformative also. It is possible that the fault may lie with the sharpness of the bound. This view is supported by the fact that L.S. estimates do surprisingly well in autocorrelated regression problems when $0 \leq \rho \leq .9$ as shown by Golub's (1963) calculations. Indeed he appears to have been required to choose suspiciously high correlations to demonstrate superiority of W.L.S. over L.S. As we have already noted, W.L.S. with weights $W = \text{diag}\,(1/v_{ii})$ is not a feasible alternative in our case, and correlations will likely be low. Although our suspicion is that L.S. will often be adequate in unbalanced cases as well as balanced$_2$ ones we are forced to conclude that there seems to be no optimal general theory covering the estimation of fixed effects in a mixed model by the above approach.

We now consider another important case where a 'goo
but apparently non-optimal solution has been in use for
years. The b.i.b. design with treatments fixed, blocks
random, is an example of the model presently under consi
ation, for which a reasonable, easily obtained estimate
treatment effects was obtained long ago. See Yates (194
and Rao (1947a), for example, both of whom say the estim
is only approximate. We know that if the weights are kn
the procedure for combining estimates gives the Aitken
(1934) or M.V. estimator. The method of attack in a b.i
makes use of the structure of the covariance matrix V an
it seems, is a more helpful approach to the estimation p
lem in a mixed model than is the more general approach s
far discussed in this section. A result by Graybill and
Weeks (1959) showing that Yates' (1940) estimator is bas
on a minimal sufficient set seems to vindicate this view
Since estimates of variance components are required to s
up the combined estimator in this case, this tends to in
cate the importance of estimation of the variance compon
over estimation of linear regression parameters. This is
the reason why this thesis concerns itself mainly with es
timation of the variance components in the model in as ef
cient a way as possible. One of the problems is to disti
guish between different methods of estimation of the vari
components which give rise to different estimators.

In some cases we can do quite a good deal, but in the majority of cases, there simply is no theory to guide us. There are two choices, a) abandon the whole question as a lost cause, or b) methodically obtain numerous estimators, with a view to later comparing them. Tukey (1962) and statisticians generally appear to subscribe, not always with enough justification perhaps, to the latter alternative.

III. ON UNBIASED ESTIMATION IN A SMALL SUBSET OF VARIANCE COMPONENTS MODELS

A. Complete Sufficient Statistics for the Completely Random Variance Components Model

1. Introduction

Let the model be represented by

$$y = j_n \mu + \sum_{i=1}^{k+1} X_i \beta_i \qquad \text{(III.A.1)}$$

where y is a n x 1 vector, y_n is a vector of unit elements, X_i's are matrices of size n x p_i of known constants and

$$E(\beta_i \beta'_i) = I_i \sigma_i^2, \quad E(\beta_i \beta'_j) = 0,$$

so that V, the variance matrix of y (= E(yy')-E(y)E(y')) is

$$X_1 X'_1 \sigma_1^2 + X_2 X'_2 \sigma_2^2 + \dots + X_{k+1} X'_{k+1} \sigma_{k+1}^2 .$$

We seek (a) necessary and sufficient conditions for the existence of M.V. unbiased estimators for μ, $\sigma_1^2, \dots, \sigma_{k+1}^2$, and (b) the form of the estimators.

A theorem due to Lehmann and Scheffè (1950) states:

If $A_i(\Theta)$ (i = 1,...,k) are estimable functions of a parametric vector $\Theta = (\Theta_1, \dots, \Theta_k)'$ and a complete sufficient statistic $T = (T_1, \dots, T_k)'$ for Θ exists, then the M.V. unbiased estimators of $A_i(\Theta)$ (i=1,...,k) exist. The M.V. unbiased estimator of $A_i(\Theta)$ is the unique function of T_i's which is an unbiased estimator of $A_i(\Theta)$. In view of this correspondence between a complete sufficient statistic and a

M.V. unbiased estimator, the first step in a search for a M.V. estimator is often a search for a complete sufficient statistic. Some distributional assumption concerning the β_i's is necessary to ensure sufficiency however; the assumption most frequently used being that β_i's be distributed independently of each other and each normally about zero with variance $I\sigma_i^2$. As it turns out, this assumption, although necessary for our development, is not always necessary for estimators to have the property of best quadratic unbiasedness. Something less will do.

Koopman (1936) showed that when the density function for a vector y can be expressed in the form

$$\exp\left\{\sum_{i=1}^{m} v_i(y)\ u_i(\theta) + C(y) + D(\theta)\right\}$$

then the elements of the vector $v(y) = (v_1(y)\ v_2(y)\ldots v_m(y))'$ are jointly sufficient for the elements of the vector $u(\theta) = (u_1(\theta)\ u_2(\theta)\ldots u_m(\theta))'$. We note that all the observations form a jointly sufficient set of statistics for the set of all parameters, and therefore, of course, it is possible for the Koopman (1936) form to do no more than indicate this. In fact, it appears that less than careful use may lead the form to indicate $n+r$ $(r > 0)$ statistics when in fact there are only n observations. The concept of a minimal sufficient set, i.e., the smallest number of sufficient statistics, could then be used to ensure that we never state that the number of sufficient statistics exceeds the number of observations, and that frequently the size of the minimal set will be substantially less than n in number.

Under the assumption of normality for the model (III. A.1), the density function for y is

$$f(y;\theta) = 1/2(2\pi)^{n/2}.|V|^{-\frac{1}{2}} \exp -\{(y-\mu)'V^{-1}(y-\mu)\}/2$$

and Graybill and Hultquist (1961), have exhibited a minimal set of sufficient statistics for this case under the further condition $X_iX'_i\ X_jX'_j = X_jX'_j\ X_iX'_i$ $(i,j = 0,\ldots,k+1)$. Since (as we shall see) commutativity of this type implies existence of an orthogonal matrix P that diagonalizes V independently of the parameters, in this case the terms i.e., $u_i(\theta)$, entering the Koopman (1936) form above, are the reciprocals of the distinct latent roots of V. If the minimal set agrees in number with the number of parameters i.e., satisfies an appropriate condition of uniqueness, or more formally if a set of sufficient statistics is also complete, we can find U.M.V. unbiased estimators for all the parameters. One convenient characterization of uniqueness in the present situation is given by the number of distinct latent roots of $E(yy') = W = X_0X'_0\mu^2 + V$. We note that although at first sight it would appear that the two conditions, (a) on commutativity of the matrices $X_iX'_i$ and (b) on the number of distinct roots of W, are non-overlapping, this is not the case. In fact if the first condition does not hold, it is not possible to diagonalize V (or W) independently of the parameters and so the latent roots of V (or W) in this case are not quantities that can be readily obtained and distinguished. Thus when we make a statement about the number of distinct latent roots of W we shall in fact imply that the condition $A_iA_j = A_jA_i$ $(i,j=0,\ldots,k+1)$ does in fact also hold. Theorem 6 of Graybill and Hultquist (1961) appears to take a different view.

Graybill and Hultquist (1961) make occasional use of a condition that the X_i matrices satisfy

$$j'_n X_i = r_i j_p{}'_i \quad \text{and} \quad X_i j_{p_i} = j_n \qquad \text{(III.A.2)}$$

where r_i is a positive integer and the subscripts n and p_i are the dimensions of respective vectors of unit elements. The sense in which this condition is required is of some interest. It turns out that mathematically there is no need for the condition in models of type (III.A.1). However, before we abandon the condition entirely, it might be as well to note that by a "variance component model" we understand something more than (III.A.1), and that, succinctly, the condition (III.A.2) brings (III.A.1) into the realm of a variance component model. Consequently when we "remove" this condition in later theorems we do so only to emphasize the conditions that are truly relevant to the particular argument at hand.

It is noteworthy that there is a mathematical (as opposed to an interpretive) need of the condition (III.A.2) in section C.

In the course of what follows, we shall restate some of the theorems proved by Graybill and Hultquist (1961) except that the condition (III.A.2) has been removed from the hypothesis in all cases where they required it. When no improvement beyond removal of the requirement (III.A.2) can be made to their proof, we shall sometimes give their proof, and indicate that the theorem is really theirs.

2. On the necessity for the commutativity condition

There is a theorem in matrix algebra which states:

Theorem 1 (well-known): For every real symmetric matrix A there is a real orthogonal matrix T such that $T'AT$ is in diagonal form.

In the model (III.A.1), where $V = \sum_{i=1}^{k+1} A_i\sigma_i^2$ and where σ_i^2 are unknown, mere symmetry of V does not suffice to allow diagonalization of V so long as V involves unknown variance components. We have already noted that the condition $A_iA_j = A_jA_i$ $(i,j=0,\ldots,k+1)$ has been shown to suffice. Let us now convince ourselves that nothing less will do.

Lemma 2 (well-known): A real symmetric matrix A is non-negative definite if there exists a matrix Q such that $A = QQ'$.

Theorem 2: If $W = X_0X'_0\mu^2 + X_1X'_1\sigma_1^2 + \ldots + I\sigma^2$ and $PWP' = \Delta$ (diagonal), independently of the parameters, and where P is orthogonal, then $PX_iX_i'P'$ $(i=0,\ldots,k+1)$ are all diagonal forms.

Proof: Since parameter values are arbitrary we may put all parameters but the first (say) equal to 0. It follows that $PX_0X'_0P'$ has to be diagonal. By repeating the argument and putting all parameters but the second equal to zero, all but the third equal to zero, and so on, we see that every form $PX_iX'_iP'$ $(i=0,\ldots,k+1)$ is in fact a diagonal one.

In other words P must simultaneously diagonalize $X_0X'_0, X_1X'_1,\ldots,X_kX'_k$. In accordance with lemma 2, and the choice of $X_{k+1} = I$ the latent roots will all be positive. Notice that $(X_iX'_i)' = X_iX'_i$, so that all these matrices are symmetric. We can therefore deal with the individual components of V (or W) as though they were a separate collection of matrices requiring simultaneous diagonalization, and this is best done by means of the following theorem, which gives a necessary and sufficient condition in terms of the X_i matrices for the existence of an adequate P matrix.

Theorem 3: (Well-known) Let $A_0 = X_0X'_0$, $A_1 = X_1X'_1,\ldots,A_t = X_tX'_t$ be a collection of symmetric n x n matrices. A necessary and sufficient condition that there exist an orthogonal transformation P such that PA_0P', $PA_1P',\ldots,PA_tP'$ are all diagonal is that A_iA_j be symmetric for all i and j. Since all A_i are symmetric, it follows that A_iA_j is symmetric if and only if A_i and A_j commute.

So much then for the need of the commutativity condition insofar as M.V. unbiased estimators in model (III.A.1) are concerned. In models of this type, V matrices will either be simultaneously diagonalizable (in which cases the commutativity condition will hold) or this will not be the case. In order to split off the small class (we shall soon see how restrictive the condition is) from the totality, we have chosen to designate the class within models (III.A.1) having $A_iA_j = A_jA_i$ $(i,j = 0,\ldots,k+1)$ by the name balanced$_2$, while all other cases are considered to be in the unbalanced class.

3. Some related theorems of interest concerning the regression parameter

We now reinforce just how restricted the class of balanced$_2$ situations is. We have seen that commutativity of A_i and A_j $(i,j=0,\ldots,k+1)$ implies that

$$X_iX'_iV = VX_iX'_i \quad (i = 0,\ldots,k+1).$$

In particular in the present model with $X_0X'_0 = A_0 = J$ we have

$$JV = VJ\ . \qquad \text{(III.A.3)}$$

This relation implies that the sum of terms in every row of V is the same and equal to the sum of terms in every column and this is therefore a necessary condition for commutativity. That the condition is not sufficient can be shown by an example. The same result is true for V^{-1}.

The next point of some interest is to determine the effects of the commutativity condition on estimates of the fixed regression parameter in the model. Since we shall later be interested in mixed models it is convenient to prove the theorems that follow for the model

$$y = X\gamma + \sum_{i=r+1}^{k+1} X_i \beta_i \qquad \text{(III.A.4)}$$

where the fixed parameters are represented collectively by γ. Of course (III.A.1) is (III.A.4) with $r = 0$,

$$X = j_n, \text{ and } \gamma = \mu .$$

Now in view of the restriction $JV = VJ$ we assert that there exists a P such that $V^{-1}j_n = j_nP$ where P is a scalar which is equal to the constant row sum of V^{-1}. In fact the properties of the least squares (L.S.) estimator for μ in the model (III.A.1) are described as a special case of the following theorem.

Theorem 4: If $X'X$ is non-singular and there exists a non-singular P such that $V^{-1}X = XP$, then $\hat{\gamma}_{M.V.} = \hat{\gamma}_{L.S.}$

Proof:
$$\begin{aligned}\hat{\gamma}_{M.V.} &= (X'V^{-1}X)^{-1}\, X'V^{-1}y \\ &= (P'X'X)^{-1}\, P'X'y \\ &= (X'X)^{-1}(P')^{-1}\, P'X'y \\ &= (X'X)^{-1}X'y \\ &= \hat{\gamma}_{L.S.}\end{aligned}$$

Theorem 5 (Zyskind et al. 1964): If $X'X$ is singular, and there exists a matrix P such that $V^{-1}X = XP$, then the estimate of any estimable function of parameters from generalized least squares equations equals the estimate of the same estimable function of parameters obtained from simple least squares equations.

Proof: The b.l.u. estimator of any estimable $\lambda'\gamma$ is unique and is of the form $\delta' X' V^{-1} y$, where δ is any solution of $X' V^{-1} X \delta = \lambda$. Since $V^{-1}X = XP$, it follows that

$$\text{b.l.u. estimator of } \lambda'\gamma = \delta' X' V^{-1} y = \delta' P' X' y$$
$$= (XP\delta)' y \;.$$

Hence the b.l.u. estimator of $\lambda'\gamma$ has for the transpose of coefficient vector the vector $XP\delta$, a vector belonging to the column space of X. Since the simple least squares estimator is the unique unbiased estimator with transpose of the coefficient vector in the column space of X, it follows that the b.l.u. estimator and least square estimators of $\lambda'\gamma$ are identical.

We note that when $X'X$ is singular the ratio $\det \Sigma_{\hat{\gamma}MV} / \det \Sigma_{\hat{\gamma}LS}$ is indeterminate.

The requirement in theorem 4 was that there exist a non-singular matrix P such that $V^{-1}X = XP$, which the reader will be aware we have not shown to be a general consequence of the commutativity condition. The following two theorems connect the condition of commutativity and the condition there exists P such that $V^{-1}X = XP$.

Theorem 6: If $VXX' = XX'V$, then the estimate of any estimable function of parameters from generalized least squares equations equals the estimate of the same estimable function of parameters obtained from simple least squares equations.

Proof: If $VXX' = XX'V$ and both V and XX' are symmetric it follows there exists an orthogonal O matrix such that both XX' and V are diagonalized simultaneously by it.

We have $O'XX'O = \Delta$ (diagonal) where (say) the 1st r terms of Δ are non-zero, and all elements of $(n-r)$ rows of $O'X$ must be zero. This implies that O' has $n-r$ rows that are orthogonal to X, i.e. O has exactly r rows in the space of X, and $(n-r)$ rows in the space orthogonal to X. The condition of existence of an O of this type that diagonalizes V is shown by Zyskind et al. (1964) to be equivalent to the condition there exists a P such that $VX = XP$, or $V^{-1}X = XP$. By Theorem 5 the result follows.

In the case where $X'X$ is singular, the ratio $\det\Sigma_{\hat{\delta}_{M.V.}}/\det\Sigma_{\hat{\delta}_{L.S.}}$ is indeterminate.

Theorem 7: If there exists P such that $V^{-1}X = XP$ and $P = P'$, then $VXX' = XX'V$.

Proof:
$$V^{-1}X = XP$$
$$\text{i.e.,} \quad X = VXP$$
$$\text{and} \quad X' = P'X'V' = PX'V\ .$$

Then
$$VXX' = VVXPPX'V$$
$$= VVV^{-1}XPX'V$$

$$= VV^{-1}XX'V$$

$$= XX'V \ .$$

The main features emerging for the model (III.A.4) are:

(a) If $X'X$ is singular, then existence of a P such that $V^{-1}X = XP$ is sufficient for $\delta'X'V^{-1}y = \rho'X'y$.

(b) Existence of a symmetric P such that $V^{-1}X = XP$ is equivalent to $VXX' = XX'V$.

These results are adequate for our purposes in later sections.

4. Restatement of known and useful theorems

We turn our attention in the remainder of this section to "weakening" the requirements for a number of theorems proved in a paper by Graybill and Hultquist (1961) dealing with the estimators of the variance components in the model (III.A.1). We do not claim that our viewpoint is greatly different from that of Graybill and Hultquist (1961), but the differences do seem to us to be worth recording in view of their clarificatory value. It is not our intention to detract from their presentation in any way, but to improve upon it.

Theorem 8: (Graybill and Hultquist (1961)) A necessary and sufficient condition that the σ_s^2 are estimable is that the A_s are linearly independent. The source of this theorem provides an adequate proof. Significantly the condition (III.A.2) is not used. This condition is however mentioned unnecessarily in the hypothesis of the following theorem, while the reference to estimability is inadvertently ignored.

Theorem 9 (Graybill and Hultquist (1961)): If in the model (III.A.1) all σ_i^2 are estimable, then the number of distinct characteristic roots of $W = E(yy')$ is not less than $k + 2$.

Theorem 10 (Graybill and Hultquist (1961)): If the number of distinct characteristic roots of W is k+2 and all σ_i^2 are estimable then the distinct characteristic roots $d_1,\ldots,d_{k+2}$ are functionally independent. (We have inserted the hypothesis that σ_i^2 are estimable, since it is implied in their proof.)

A clarification of the relation between distinct latent roots of V and of W may be helpful at this time. We have $W = E(yy') = \sum_{i=0}^{k+1} A_i\sigma_i^2$; $V = E(yy') - E(y)E(y') = \sum_{i=1}^{k+1} A_i\sigma_i^2$.

If V is diagonalized by $P = (P_1'\ldots P_s')'$, where P_i is a matrix made up of a complete collection of independent latent vectors corresponding to the i^{th} distinct latent root of V, then

$$PVP' = \sum_{i=1}^{k+1} \sigma_i^2 D_i \quad\text{and}\quad PWP' = PA_0P'\sigma_0^2 + \sum_{i=1}^{k+1} \sigma_i^2 D_i .$$

P can be chosen so that $PVP' = \mathrm{diag}(d_1^*, d_2I_2,\ldots,d_sI_s)$ where $d_1^* = d_1 - n\mu^2$ and $d_1,\ldots,d_s$, are the s distinct roots of W, i.e.,

$$PWP' = \begin{bmatrix} n\sigma_0^2 & & \\ & \ddots & \\ & & 0 \end{bmatrix} + \sum_{i=1}^{k+1} \sigma_i^2 D_i = \mathrm{diag}(d_1, d_2I_2,\ldots,d_sI_s).$$

In consequence, all latent roots of V and W except for the first are in agreement. If the 1st root of V is in agreement with some other root then W will have one more distinct root than V; if on the other hand the first root is not identical with any other root, V and W will have the same number of distinct roots. The more important consideration for our purposes, is the number of distinct roots of W. From what has been said above this number will be the same, or one more than the number of distinct roots of V. Graybill and Hultquist (1961) state and prove only part of theorem 11 given below, i.e., if W has k+2 distinct roots then $y'P'_iP_iy$ ($i=2,\ldots,k+2$) and P_1y form a complete sufficient set. The proof of this part is essentially theirs. Insertion of the commutativity condition may well strike the reader as surprising, in view of the work of Imhof (1960) on a mixed model of Scheffè (1956, 1959). Imhof finds a complete sufficient set of statistics for parameters of a highly restricted cross-classification with equal numbers model, which is not within the class (III.A.1), and hence not covered by our classification, without requiring that V be diagonalizable.

Theorem 11: In a completely random model (III.A.1) assuming commutativity and normality, and all σ_i^2's estimable, there is a complete sufficient set of statistics for the parameters $\mu, \sigma_1^2,\ldots,\sigma_{k+1}^2$ if, and only if, W has k+2 distinct latent roots.

Proof: Consider the joint distribution of $y_1,\ldots,y_n$. The quadratic form

$$Q = (y - \mu)' V^{-1} (y - \mu)$$

$$= (P(y-\mu))' PV^{-1}P' (P(y-\mu))$$

can be written

$$Q = d_1^{*-1} (P_1 y - n^{\frac{1}{2}} \mu)^2 + \sum_{u=2}^{s} d_u^{-1} y'P'_u P_u y$$

where $s = k+2$.

In view of theorem 10 above, the k+2 distinct characteristic roots $d_1, d_2, \ldots, d_s$ are functionally independent. The Neyman factorization theorem gives $T = (P_1'y, y'P_2'P_2y, \ldots y'P_s'P_sy)$ for the set of sufficient statistics. It is necessary to make use of a lemma by Gautschi (1959) and write the joint density for T in a form that involves a product of two independent distributions whence we can infer that the set of statistics $T = (P_1y, y'P_2'P_2y, \ldots y'P_s'P_sy)$ which is the estimator for $(\mu, \sigma_1^2, \ldots, \sigma_{k+1}^2)$, is indeed complete.

Conversely, we now show that if under commutativity and normality there exists a complete sufficient set of statistics for the parameters $(\mu, \sigma_1^2, \ldots, \sigma_{k+1}^2)$ then W has k+2 roots. In view of commutativity, we have an orthogonal A.o.V. breakdown. In fact if there exists a P such that $PWP' = \Delta$ (diagonal) then we have $Py = (P_1'P_2' \ldots P_s)'y$ (say) where P_i is $m_i \times n$ (say) and where m_i is the multiplicity of the root δ_i. Let there be s distinct roots. Since all σ_i^2's are estimable $s \geq k+2$. It follows that the A.o.V. breakdown is

$$y'y = \sum_{i=1}^{s} y'P_i'P_i y$$

where $E(y'P_i'P_i y) = c_i \delta_i$. Ignore for the moment the component $y'P'_1P_1y$ that corresponds to the mean. Completeness implies that there are exactly k+1 other components, i.e. $s = k+2$. This follows since if $g(\theta)$ is an unbiased estimator for which unbiased estimators $S_1(T) = S_1$ and $S_2(T) = S_2$ depending on a complete sufficient statistic T can be found, then $S_1 = S_2$. Consequently $y'y = \sum_{i=1}^{k+2} y'P_iP_iy$ and W has k+2 distinct roots.

In view of the Lehmann-Scheffè (1950) theorem we can infer that if W has k+2 distinct roots and all σ_i^2's are estimable, then the standard A.o.V. gives M.V. unbiased estimators for unique functions of σ_i^2's.

The real problem arises when W has more than k+2 distinct roots or commutativity does not hold. In both these cases we are for all practical purposes in difficulty.

Lemma 3 (Huzurbazar (1963)): When an "efficient" or minimum variance unbiased estimator exists it is unique. Rao (1945) showed that the M.V. unbiased estimator of a parametric function $\tau(\theta)$, when it exists, is always a function of the sufficient statistic t. The generalization of this to k parameters to which we shall appeal was given by Rao (1947b):

for the simultaneous estimation of $r(\leq k)$ functions τ_i of the k parameters θ_s, the unbiased functions of a minimal set of k sufficient statistics, say t_i, have minimal attainable variances. Hence the search for the M.V. unbiased estimator of τ_i's ($i=1,\ldots r$, $r\leq k$) reduces to consideration of functions of t_i's ($i=1,\ldots,k$) which are unbiased estimators of τ_i's. If there exist several such functions of τ_i's which are unbiased for t_i then the function of such several functions with least variance is the M.V. unbiased estimator of τ_i. In view of lemma 3, such a function of a sufficient statistic with least variance will be unique, since the M.V. unbiased estimator is unique when it exists.

We feel that the following definition is of some value in distinguishing different cases in the model (III.A.1): All cases in the model (III.A.1) fall into one of the three classes P,R-P, S-R where

$$P \subset R \subset S$$

and C means "is a proper subset of". P is the class of situations in which W is diagonalizable and has a distinct number of roots equal to the number of parameters, R is the class of situations where W is diagonalizable and S is the unrestricted class of situations. Let us redefine the matrix P that diagonalizes W as follows: $P = (P_0' \ldots P_{k+1}')'$.

In later theorems on a mixed model we need the condition $P_i X_j = 0$ $(i\neq 0) \neq (j\neq k+1)$. One consequence of this condition, which in effect defines a sub-class of P, is that, apart from roots corresponding to P_0, the non-zero roots of individual A_i matrices making up W are added in non-overlapping places. In this sub-class of P, the breakdown is a unique one in terms of orthogonal projection operators on orthogonal complements of the vector j_n in the column spaces of the X_i matrices.

The properties of A.o.V. estimators in situations described by class P have been shown to be M.V. by Theorem 11.

The situation in the other two classes i.e. R-P and S-R is not clear-cut and is debatable. We take this up later. It will suffice at this time to point out that in class R a minimal set of sufficient statistics does exist, but we do not have any machinery for constructing an M.V. estimator in that case. In fact it is not clear on what grounds Rao's (1947a) conjecture, that there is a U.M.V. estimator in a special example of this nature, is based.

We conclude this sub-section with a theorem which indicates that the good properties for estimators derived under an assumption of normality for class P above carry over when less restrictive assumptions instead of normality are made. Theorem 12 (below) is our interpretation of theorem 7 of Graybill and Hultquist (1961).

We note that although the assumption of normality is removed, independence of vectors β_i and β_j $(i \neq j)$ and between elements within a given vector β_i, is not. This means, for example, that terms of type $E(\beta_{i1}{}^3\beta_{i2})$ are all assumed to be zero. The justification for this assumption is that we are sampling from an infinite population.

Theorem 12: In the model (III.A.1) consider only the class P. If β_i and β_j are independent for all i and j ($i \neq j$) and finite fourth moments exist for all random variables, and within every given vector β_i, all fourth moments are equal, and all third moments are equal, then the same estimators, i.e. the usual Model I. A.o.V. mean square estimators for the $\delta_i = E(y'P'_iP'_iy)$, that are M.V. unbiased under normality, are b.q.u. estimators under present assumptions.

The proof of this result given by Graybill and Hultquist (1961) is an abbreviated one, and is in fact obscure. Since we shall state and prove a more general version of this theorem in section C, we do not expand on this proof at this time. We note that Theorem 12 is a considerable generalization of a result proved by Graybill (1954) for b.q.u. estimators under similar restrictions in the special case of a general "balanced" nested design.

We conclude this section by noting that:

(a) we have purposely done away with the restricted definition of existence of an A.o.V. made by Graybill and Hultquist (1961) because we do not favor defining an A.o.V.

to exist when and only when the number of sums of squares in the subdivision agrees with the number of parameters to be estimated.

(b) We have made no restriction to classification type models since none is needed. In the event a non-classification type model can be found which satisfies the commutativity condition and has distinct roots equal in number to the number of parameters, then M.V. properties can be inferred for estimators "borrowed" from Model I. No natural non-classification examples are presently known.

B. On an Extension to a Mixed Model

In brief our objectives in this section are to show that the same results, namely a complete sufficient set of statistics and hence M.V. estimators under normality, and b.q.u. estimators when normality is replaced by slightly less stringent conditions, can be obtained for certain mixed model situations. We find it convenient to fix ideas and indicate procedures by means of the best known example of a "mixed" model (although not usually considered as such) namely the general linear hypothesis model (i.e., Model I.).

$$y = X\gamma + e$$

where $y(n \times 1)$ is a vector of observations,

$X(n \times p)$ is a known matrix of rank r of constants,

γ (p x 1) is a vector of unknown parameters

e(n x 1) is a vector of normally distributed error elements satisfying $E(e) = 0$, $E(ee') = I\sigma^2$.

The content of Theorem 13 (to follow) is essentially known, but no proof appears to have been given. In the proof of Theorem 13 we shall use the following:

Lemma 4 (Imhof (1960)): Let θ be a parameter vector and Y be a random vector in Euclidean space E_n, similarly let θ_1 and Y_1 be vectors in E_{n1}. Assume that Y and Y_1 have probability densities (with respect to Lebesque measure) of the form

$$p(Y,\theta) = g(\theta)\ h(Y)\ \exp\{\theta' Y\}$$
$$p(Y_1,\theta_1,\theta) = f(\theta_1,\theta)\ \exp\left\{ Y'_1 R(\theta)\ Y1 + \theta'_1 Y_1 \right\}$$

where $R(\theta)$ is a matrix of size n_1. Let the domain $\mathfrak{D}$ of θ contain a non-degenerate interval in E_n and the domain of θ_1 be E_{n1}. Then, the family of product measures on E_{n+n1} generated by the family of probability densities

$$T = \left\{ p(Y,\theta)\ p_1(Y_1,\theta_1,\theta)\ :\ (\theta,\theta_1)\varepsilon\, \mathfrak{D} \times E_n \right\}$$

is strongly complete in the sense of Lehmann and Scheffè (1955).

Let $\hat{\gamma}$ be any vector satisfying the normal equations $X'X\gamma = X'y$ and $(n-r)s^2 = (y-X\hat{\gamma})'(y-X\hat{\gamma}) = y'y-y'X(X'X)^*X'y$ where $(X'X)^*$ is any matrix satisfying $X'X(X'X)^*X'X=X'X$, i.e., $(X'X)^*$ is a conditional inverse of $X'X$.

Theorem 13: In the general linear model $y = X\gamma + e$, the statistic $(X\hat{\gamma}, s^2)$ is both sufficient and complete for the parameter set $(X\gamma, \sigma^2)$.

Proof: The quadratic form in the exponent of the density function for y, ignoring the $-\frac{1}{2}$, is

$$Q = (y - X\gamma)'(I\sigma^2)^{-1}(y - X\gamma)$$

$$= (y - X\hat{\gamma})'(y - X\hat{\gamma})/\sigma^2 + (X\hat{\gamma} - X\gamma)'(X\hat{\gamma} - X\gamma)/\sigma^2$$

$$= (n - r)s^2/\sigma^2 + (X\hat{\gamma} - X\gamma)'(X\hat{\gamma} - X\gamma)/\sigma^2 .$$

Sufficiency follows by the Neyman factorization theorem. Now $X\hat{\gamma}$ and s^2 are independent, s^2/σ^2 is $c\chi^2_{n-p}$, and $X\hat{\gamma}$ is $N(X\gamma, X(X'X)^*X'\sigma^2)$.

If we make the following correspondence with terms in Lemma 4, namely

$$s^2 = Y, \qquad \sigma^2 = \theta,$$

$$X\hat{\gamma} = Y_1, \text{ and } X\gamma = \theta_1 \qquad \text{then}$$

if $T = (X\hat{\gamma}, s^2)$ it follows that the density of T becomes

$$p_T = p(Y,\theta)\, p_1(Y_1,\theta_1,\theta)$$

and the family of product measures generated by p_T is therefore strongly complete.

The statistic $T = (X\hat{\gamma}, s^2)$ is thus complete sufficient for the set $(X\gamma, \sigma^2)$.

We turn now to the conventional mixed model

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=r+1}^{k+1} X_i \beta_i \qquad \text{(III.B.1)}$$

which we shall also represent by

$$y = X\gamma + \sum_{i=r+1}^{k+1} X_i \beta_i \qquad \text{(III.B.1a)}$$

where γ_i's and γ are fixed parameters and β_i's are statistically independent vectors each distributed according to the multivariate normal distribution $N(0, I\sigma_i^2)$, $i = r+1, \ldots, k+1$. We shall say that the mixed model (III.B.1) satisfies the balance$_2$ condition if

$$X_i X_i' X_j X_j' = X_j X_j' X_i X_i' \quad (i,j=0,\ldots,k+1) \quad .$$

In this section we restrict ourselves to those mixed models that satisfy the balance$_2$ condition. Notice that the restrictions for balance$_2$ in the representations (III.B.1) and (III.B.1a) are not identical.

The linear model representation $y = X\gamma + e$ with $E(e) = 0$, and $E(ee') = I\sigma^2$ discussed previously satisfies $XX'I = IXX'$ and is therefore balanced$_2$. We specialize (III.B.1) further to have $A_0 = J$, $A_{k+1} = I$. Then $V = \sum_{i=r+1}^{k+1} A_i\sigma_i^2$; $E(yy') = W = X\gamma\gamma'X' + V$. If we choose the representation

$$y = j_n\mu + \sum_{i=1}^{k+1} X_i \beta_i \qquad \text{(III.B.2)}$$

as the corresponding completely random model (for(III.B.1)), y in (III.B.2) has variance $\overline{V} = \sum_{i=1}^{k+1} A_i\sigma_i^2$, and if the commutativity condition is satisfied we can find an orthogonal matrix P such that

$$P\overline{V}P' = \Delta \text{ (diagonal)}$$
$$= PA_1P'\sigma_1^2 + PA_2P'\sigma_2^2 + \ldots + PP'\sigma_{k+1}^2$$

and we denote diagonal elements of Δ by δ_i's .

Furthermore we have

$$PVP' = \Delta^* \text{ (diagonal)}$$
$$= PA_{r+1}P' \ \sigma_{r+1}^2 + \ldots + P'P \ \sigma_{k+1}^2$$

and we denote diagonal elements of Δ^* by δ_i^*'s. Now if $\overline{W}$ has k+2 distinct roots, and if furthermore $P_iX_j = 0$ $(i \neq 0) \neq (j \neq k+1)$ then it is easily seen by the argument following theorem 11 that V has exactly k-r+1 distinct roots. We have chosen to designate these as δ_i^*'s. We have been unable to proceed to complete sufficient statistics without at least $P_iX_j = 0$ $(i \neq 0,\ r+1 \leq j \leq k)$ and V has exactly k-r+1 distinct roots. The assumption $P_iX_j = 0$ $(i \neq 0) \neq (j \neq k+1)$ is sufficient for the latter ; it is not claimed to be necessary.

If we assume normality for the β_i's, in (III.B.1) and write out the density of the observation vector y, then we obtain the quadratic form of the exponent ignoring the $-\frac{1}{2}$, to be

$$Q = (y-X\gamma)'\ V^{-1}(y-X\gamma) = (P(y-X\gamma))'(PV^{-1}P')(P(y-X\gamma))$$
$$= (P_0(y-X\gamma))'(P_0(y-X\gamma))/\delta_h^* + (P_1(y-X\gamma))'(P_1(y-X\gamma))/\delta_j^*$$
$$+ \ldots + (P_{k+1}(y-X\gamma))'(P_{k+1}(y-X\gamma))/\delta_m^* \qquad \text{(III.B.3)}$$

where the P_i matrices are defined as subspaces of P with dimensions equal to the multiplicities of the roots of $\overline{V}$. These subspaces are made up of vectors corresponding to similar latent roots or δ_i's i.e. $P = (P_0'P_1'P_2' \ldots P_r'P_{r+1}' \ldots P_{k+1})'$ and there are k+2 in all; it follows that in (III. B.3) δ_i^* values do not occur only once. Nor can we specify which δ_i^* goes with any particular term. The restriction $P_iX_j = 0$ $(i \neq 0) \neq (j \neq k+1)$ simplifies the expression (III.B.3) to $(P_0y-P_0X\gamma)'(P_0y-P_0X\gamma)/\delta_h^* + (P_1y-P_1X\gamma)'(P_1y-P_1X\gamma)/\delta_j^*$

$$+ \ldots (P_ry-P_rX\gamma)'(P_ry-P_rX\gamma)/\delta_k^* + \sum_{i=r+1}^{k+1} y'P_i'P_iy/\delta_i^*$$

$$= (R-ER)' R_0^{-1}(R-ER) + \sum_{i=r+1}^{k+1} y'P_i'P_iy/\delta_i^*$$

where $R = \overline{P}y$, $ER = \overline{P}X\gamma$

R_0^{-1} = a diagonal matrix of δ_i^*'s, and $P = (P_0'P_1' \ldots P_r')'$.

In view of the Neyman criterion and prior knowledge, namely that $P_i(P_i'P_i)^*P_i'$ is a projection on the space of X_i and P_iy $(i \leq r)$ is therefore the exact linear function that is indicated by the usual Model I breakdown or L.S. procedure, and $X\hat{\gamma}_{LS} = X\hat{\gamma}_{MV}$, we conclude that $(X\hat{\gamma}_{LS}, s_{r+1}^2, \ldots s_{k+1}^2)$ is sufficient for the parameter set $(X\gamma\ , \sigma_{r+1}^2, \ldots \sigma_{k+1}^2)$.

Theorem 14: The condition of balance$_2$ and $P_iX_j = 0$ $(i \neq 1) \neq (j \neq k+1)$ in a mixed model $y = X\gamma + \Sigma X_i\beta_i$ where β_i's are independent normal and where there are k+2 roots for $\overline{W}$ in the corresponding completely random model (III.B.2) ensures that the statistic $T = (X\hat{\gamma}_{LS}, s_{r+1}^2, \ldots, s_{k+1}^2)$ is complete sufficient for the parameters of the model.

Alternatively: If the Model (III.B.1) satisfies the restrictions
$XX'V = VXX'$, $P_jX = 0$ $(j=r+1,\ldots,k)$, $X_iX_i'X_jX_j' = X_jX_j'X_iX_i'$ $(i,j=r+1,\ldots,k+1)$, V has k-r+1 distinct roots and β_i's are independent normal then the statistic $T = (X\hat{\gamma}_{LS}, s_{r+1}^2, \ldots, s_{k+1}^2)$ is complete sufficient for the parameters of the model.

Proof: We have that $(y - X\hat{\gamma}_{LS})$ is distributed $N(0,V)$. Furthermore,

$$P_i'P_iV/\delta_i^* \cdot P_i'P_iV/\delta_i^* = P_i'P_iV/\delta_i^* \text{ and}$$

$$P_i'P_iV/\delta_i^* \; P_j'P_jV/\delta_j^* = 0 \,,$$

in view of the choice of P_i, P_j as subspaces of an orthogonal matrix P. It follows that s_i^2's are independently distributed as $c_{p_i}\chi_i^2$. In a balanced$_2$ case $X\hat{\gamma}_{LS}$ is multivariate normal $(X\gamma, R_0)$.

Since P_iy $(0 \leq i \leq r)$ and P_iy $(r+1 \leq i \leq k+1)$ are uncorrelated, in view of normality we have independence of all statistics actually calculated. If we make the following correspondence between present variables and parameters and those in lemma 4

$$(s^2_{r+1},\ldots,s^2_{k+1})' = Y, \quad (\sigma^2_{r+1},\ldots,\sigma^2_{k+1})' = \theta ,$$

$$Py = Y_1 \text{ and } \bar{P}X\gamma R_0^{-1} = \theta_1$$

$$\text{and } R(\theta) = R_0^{-1}$$

then it can be verified that

$$P_T = P(Y, \theta)\, P_1(Y_1, \theta_1, \theta)$$

where the two factors are of the type considered by Imhof (1960) and the family of product measures generated by P_T is therefore strongly complete. We conclude that the sufficient statistic $T = (X\hat{\gamma}_{LS}, s^2_{r+1},\ldots,s^2_{k+1})$ is complete.

We conclude this section by stating without proof the mixed model version of theorem 12.

Theorem 15: Suppose that in the model (III,B.1) we consider the class P, i.e., that class where complete sufficiency can be established under normality. If finite fourth moments exist for all β_i random variables, and all third and fourth moments are equal for all variables in a given vector β_i, and independence is retained, then the same estimators for $\delta_i = E(y'P_i{}'P_iy)$ that are M.V. unbiased under normality are b.q.u. under present assumptions while the ordinary least squares estimators for regression coefficients are b.l.u. estimators.

In conclusion we note that the results of this section have relevance for example in mixed classification models without interaction of which the randomized blocks case with blocks random and treatments fixed is an example. Hultquist and Graybill (1965) exhibit a minimal set for this situation; we have shown that the estimators usually used are in fact complete. This makes a considerable difference to the properties that we can infer for the "usual" L.S. estimators.

C. Other Models

In this section we consider briefly estimation in two basic types of models. They are of special interest because of their relationship to covariance structures induced under a finite model.

The first type is represented by

$$y = \sum_{i=0}^{k+1} X_i \beta_i$$

$$X_0 = j_n , \beta_0 = \mu , X_{k+1} = I, \text{ and } E(\beta_{k+1}\beta_{k+1}') = Ia_{k+1}$$

For $i = 1,\ldots,k$, we have $E(\beta_i\beta_i') = (a_i \backslash b_i) = (a_i - b_i)I_i + b_i J_i^i$ where

$(a_i \backslash b_i)$ is a matrix with a_i's on the diagonal and b_i's off it, while for $i=1,\ldots,k+1$, $E(\beta_i \beta_j') = 0$ $(i \neq j)$. (III.C.1)

The second type is represented by

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=r+1}^{k+1} X_i \beta_i$$

$X_0 = j_n$, $\gamma_0 = \mu$, $X_{k+1} = I$ and $E(\beta_{k+1}\beta_{k+1}') = Ia_{k+1}$.

For $i = r+1,\ldots,k$, we have $E(\hat{\beta}_i \hat{\beta}_i') = (a_i \setminus b_i)$, and for $i=r+1,\ldots,k+1$, $E(\hat{\beta}_i \hat{\beta}_j') = 0 \quad (i \neq j)$. (III.C.2)

Our first objective is to exhibit M.V. unbiased estimators for some of these cases. We shall draw attention to the fact that in such cases the estimators of previous sections are still M.V. estimators, but that the estimators of variance components estimate the (a_i-b_i)'s $(b_{k+1}=0)$. Generalizations of various theorems previously given will be presented. A simple example, which exhibits the essential features of constant correlation in (III.C.1), is

$$y = \mu + e$$

where e has covariance matrix $V = (a-b)I + bJ$. Let P be any orthogonal matrix that diagonalizes V. It is known that $\bar{y}$ is a MV estimator for μ. Let us assume normality of distribution and write down the density function for y. The quadratic form in the exponent, ignoring the $-\frac{1}{2}$ is,

$$\begin{aligned} Q &= (y-\mu)' \, V^{-1} \, (y-\mu) \\ &= (P(y-\mu)'(PVP')^{-1}P(y-\mu) \\ &= (P_0(y-\mu)'P_0(y-\mu)/\{a+(n-1)b\} + (P_1(y-\mu))'P_1(y-\mu)/a-b \\ &= n(\bar{y}-\mu)^2/a+(n-1)b + y'P'_1 P_1 y/a-b \end{aligned}$$

since $P_0(y-\bar{y}) = 0$ and $P_i j_n \mu = j_n 0$. Since $P_0(y-\mu)$ and $P_1(y-\mu)$ are orthogonal, the two component parts of Q are independent. Furthermore $s^2 = y'P'_1 P_1 y/(a-b)$ has $\chi^2(n-1)$ distribution and $(\bar{y}-\mu)$ is distributed $N(0,(a+(n-1)b)/n)$.

By appealing to Gautschi's (1959) lemma we therefore have:

Theorem 16: The statistic $T = (\bar{y}, s^2)$ in the model $y=\mu+e$ where $E(ee') = V = (a\backslash b)$ is a complete sufficient statistic for $(\mu, a-b)$.

The parameter $(a+(n-1)b)$ cannot now be obtained as a function of the set $(\mu, a-b)$, as was the case in previous cases. It can also be shown that there are no unbiased estimators of the individual parameters a or b. The example demonstrates that complete sufficiency may exist for a vector function of parameters but not for the set of all the parameters themselves. A similar situation obtains with respect to unbiased estimators.

We verify that this state of affairs carries over to some more complex models. We restrict our considerations to models of type (III.C.1) which satisfy the balance$_2$ restriction

$$X_iX_i'X_jX_j' = X_jX_j'X_iX_i' \quad (i,j=0,\ldots,k+1) \qquad \text{(III.C.5)}$$

and for which the additional restriction

$$j_n'X_i = r_ij_{p_i}' \ , \ X_ij_{p_i} = j_n \qquad \text{(III.C.6)}$$

is imposed. It follows that $E(yy') = W$

$$= X_0X_0'\mu^2 + \sum_{i=1}^{k} X_i((a_i-b_i)I_i+b_iJ_i^i)X_i' + Ia_{k+1}^2$$

$$= X_0X_0'\mu^2+X_1X_1'(a_1-b_1)+\ldots+X_kX_k'(a_k-b_k)+Ia_{k+1}^2+X_1J_1^1X_1'b_1+ \ldots$$
$$+X_kJ_k^kX_k'b_k$$

$= J_n^n \mu^2 + V_1 + J_n^n(b_1 + \ldots, b_k) = V_1 + J_n^n(\mu^2 + b_1 + \ldots + b_k)$.

We note that V_1 agrees exactly with the term from the corresponding model without correlation, i.e. (III.A.1) if we replace σ_i^2 therein by $(a_i - b_i)$ $(i=1,\ldots,k)$ and σ_{k+1}^2 by a_{k+1}. We emphasize that condition (III.C.6) has been utilized. As in Theorem 11 there exists an orthogonal P such that $PWP' = \Delta$ diagonal, where elements of Δ are denoted by δ_i's.

Theorem 17: Under an assumption of a multivariate normal distribution in the model (III.C.1) which satisfies also (III.C.5) and (III.C.6) if $W = V + X_0X_0'\mu^2$ has k+2 distinct latent roots, then the vector-statistic

$T = (P_0y,\ s_1^2,\ s_2^2,\ldots,s_{k+1}^2)$ is complete sufficient

for $(\mu, (a_1-b_1),\ldots,(a_k-b_k),\ a_{k+1})$.

Proof: The quadratic form in the exponent of the density function for y, ignoring the $-\frac{1}{2}$, is

$$
\begin{aligned}
Q &= (P(y-\mu)'(PV^{-1}P')(P(y-\mu)) \\
&= (P_0(y-\mu))'(P_0(y-\mu))/\lambda_0 + (P_1y)'(P_1y)/\delta_1 + \ldots + \\
&\qquad (P_{k+1}y)'(P_{k+1}y)/\delta_{k+1} \\
&= (P_0(y-\mu))'(P_0(y-\mu))/\lambda_0 + c_1s_1^2/\delta_1 + \ldots + c_{k+1}s_{k+1}^2/\delta_{k+1}.
\end{aligned}
$$

Mutual independence of all terms $P_0y,\ s_1^2,\ldots,s_{k+1}^2$ is assured and by Gautschi's (1959) lemma the result follows.

Therefore although we cannot find estimators for all the parameters, we do find that there is a certain robustness to the goodness of the usual mean squares for estimating adjusted variance components i.e. (a_i-b_i)'s $i=1,\ldots,k+1, b_{k+1}=0$, in the model (III.C.1) with correlated effects.

The theorem (to follow) is a generalization of theorem 12. We noted earlier that the proof of theorem 12 that appears in the literature is inadequate. We are able to establish that Theorem 12 is true by observing that the nature of $P_i'P_i$ matrices in the class P is a highly restricted one which we shall describe and to which we must therefore necessarily limit our proof.

Theorem 18: For the model (III.C.1), under the additional restriction (III.C.6) and allowing only one non-zero entry per row for X_i matrices, consider the class of situations for which it is true that the submatrices B_{ii} and B_{ij} of the projection operators $P_i'P_i$ satisfy appropriate restrictions (to be specified).

If instead of normality, we have independence of β_i and β_j vectors $(i\neq j)$ and if furthermore all fourth moments exist for the elements of each β_{il} and $E(\beta_{ik}^4) = c_i$, $E(\beta_{ik}^3) = d_i$, $E(\beta_{ik}^3 \beta_{il}) = e_i$,

$E(\beta_{ik}^2\beta_{il}\beta_{im}) = f_i$, and $E(\beta_{ik}\beta_{il}\beta_{im}\beta_{in}) = g_i$ $(i=1,\ldots,k,\ \ e_{k+1}= f_{k+1} = g_{k+1} = 0$) where k, l, m and n all unequal and c_i, d_i, g_i are constants, then the estimators that are M.V.U. under normality are b.q.u. estimators.

Proof: In class P, there exists an orthogonal matrix $P = (P_0'P_1' \ldots P_{k+1}')'$ such that

$$y'y = y'P_0'P_0 y + \ldots + y'P_{k+1}'P_{k+1}\, y \; .$$

Matrices of the type $P_i'P_i$ are idempotent, symmetric and regular in structure. This will be seen to be crucial to the argument to follow.

Let the general quadratic estimator of δ_i be $\hat{\delta}_i$. Expressing this in terms of $y'P_i'P_i y$ we have

$$\hat{\delta}_i = y'P_i'P_i y + y'C^i y \qquad \text{(III.C.7)}$$

where the constant elements of C^i are defined by the relation (III.C.7). Now $E(\hat{\delta}_i) = E(y'P_i'P_i y) + E(y'C^i y)$, so that unbiasedness implies

$$E(y'C^i y) = E\left\{ \Sigma c^i_{jk} (\mu + \sum_{i=1}^{k+1} X_i \beta_i)_j (\mu + \sum_{i=1}^{k+1} X_i \beta_i)_k \right\} .$$

The vector y is made up of μ and the "contributions" $X_e\beta_e$. In view of (III.C.6) and since we allow only one entry per row for X_i matrices we notice that it is possible to rearrange $X_e\beta_e$ in the form $(\beta_{e1}, \beta_{e1}, \ldots, \beta_{e1}, \beta_{e2}, \ldots, \beta_{e2}, \ldots, \beta_{et})'$ an n x 1 vector. We explain in Chapter V that each vector β_e entering y may be regarded to denote a "β_e-partition" of the matrix of the quadratic form B (say) of the type

$$\begin{bmatrix} B_{11}\beta_{e1}\beta_{e1} & B_{12}\beta_{e1}\beta_{e2} & \cdots \\ B_{21}\beta_{e2}\beta_{e1} & B_{22}\beta_{e2}\beta_{e2} & \cdots \\ \cdot & \cdot & \cdots \end{bmatrix}$$

where B_{11} represents a certain left-hand uppermost block of the matrix B. This subdivision of B is called a β_e-partition.

$$\text{Now } 0 = E(y'C^iy) = \sum_{jk} c^i_{jk}\,\mu^2 + \sum_{\beta_e}\sum_{r} C^i_{rr(\beta_e)}E(\beta_e^2) + \sum_j c^i_{jj}\sigma^2_{k+1}$$

where $C^i_{rr(\beta_e)}$ is the sum of all c^i_{jk} terms in the rth diagonal block as determined by the β_e-partition. By equating coefficients we obtain numerous restrictions that the C^i matrix must satisfy. For the estimate $\hat{\delta}_i$ to be "best" it is required that $E(\hat{\delta}_i^2) - (E(\hat{\delta}_i))^2$, or since $E(\hat{\delta}_i) = \delta_i > 0$, more simply $E(\hat{\delta}_i^2)$ must be a minimum. We have

$$E(\hat{\delta}_i^2) = E(y'P_i'P_iy)^2 + 2E(y'P_i'P_iy)(y'C^iy) + E(y'C^iy)^2 \quad .$$

We can choose C^i to make this a minimum if the center term vanishes identically. Put $B_i = P_i'P_i$. Then we can write

$$E(y'B_iy \,.\, y'C^iy) = E\Big\{\Big(\sum_{jk} b_{jk}\mu^2 + \sum_{jk} b_{jk}\mu\,\beta_k + \sum_{jk} b_{jk}\mu\beta_j$$

$$+\sum_{jk} b_{jk}\beta_j\beta_k + \sum_{jk} b_{jk}\mu e_j + \sum_{jk} b_{jk}\mu e_k + \sum_{jk} b_{jk}e_je_k$$

$$+ \sum_{jk} b_{jk}\beta_je_k + \sum_{jk} b_{jk}e_j\beta_k\Big)$$

times a similar expansion for $y'C'y\Big\}$.

We illustrate why all terms vanish. The coefficient of

μ^4 is

$$(\sum_{jk} b_{jk})(\sum_{jk} c^{i}_{jk}) = (\sum_{jk} b_{jk})(0) = 0$$

while the coefficient of $E\,\beta_j^{\,4}$ is

$(B_{jj}\ (\beta_j) \otimes C_{jj}\ (\beta_j)\ )$ where $\otimes$ means Kronecker product of matrices. Equality of fourth moments implies that the coefficient of $E\,\beta_j^{\,4}$ is

$$\sum_j (B_{jj}\ (\beta_j) \otimes C_{jj}\ (\beta_j)) \ .$$

It is certainly not obvious why this quantity is always zero. In fact this is only the case because of the highly restricted nature of the $P_i{}'P_i$ matrices. Depending on the vector that determines the partition, different sub-matrices of $P_i{}'P_i$ are to be regarded as B_{jj}'s and B_{ij}'s. The present theorem limits itself to those cases in which with respect to the relevant partitions in each case, the matrices B_{jj} are in fact equal and the matrices B_{ij} are equal but for a possibly different sign. We therefore have

$$\sum_j B_{jj}(\beta_j) \otimes C_{jj}(\beta_j) = B_{jj}(\beta_j) \otimes \sum_j C_{jj}(\beta_j) = 0 \ .$$

Terms in $E\,\beta_i^{\,3}\beta_j$ (in abbreviated notation) are

$$\sum_{ij} B_{ii} \otimes C_{ij} + \sum_{ij} B_{ij} \otimes C_{ii}$$

$$= B_{ii} \otimes \sum_{ij} C_{ij} + \sum_j B_{ij} \otimes \sum_i C_{ii} = 0 \ .$$

We claim that the restrictions given ensure that all terms of $E(y'P_i'P_iyy'C^iy)$ do in fact vanish, and

$$E(\hat{\delta}_i)^2 = E(y'P_i'P_iy)^2 + E(y'C^iy)^2 \quad .$$

$E(\hat{\delta}_i)^2$ takes on its minimum value when

$$E(y'C^iy)^2 = 0.$$

Now $E(y'C^iy) = E(y'C^iy)^2 = 0$ implies $C^i = 0$ is the best choice we can make of C^i and the best quadratic unbiased estimate of δ_i (some linear function $\sigma_i{}^2$s) is $y'B_iy$.

In model (III.C.2), as in (III.C.1) we assume conditions (III.C.5) and (III.C.6). It follows that

$$E(yy')-E(y)E(y') = V = V_1+J_n^n(b_{r+1}+b_{r+2}+\ldots+b_k)$$

where V_1 is the variance term for the corresponding mixed model (III.B.1) with $E(\beta_i\beta_i') = I\,\sigma_b^2$ considered in section B, if we replace σ_i^2 therein by a_i-b_i $(i=r+1,\ldots,k)$, and σ_{k+1}^2 by a_{k+1}.

The corresponding model with all factors random is (III.C.1). We recall however that when the model (III.C.2) is being considered, for convenience we designate the covariance matrix of (III.C.1) by $\overline{V}$. There exists an orthogonal P such that $P\overline{V}P' = \Delta$ diagonal. Suppose $\overline{W} = \overline{V} + A_0\mu^2$ has k+2 distinct roots. It follows that $PVP' = \Delta^*$ diagonal and if furthermore $P_iX_j = 0$ $(i\neq 0) \neq (j\neq k+1)$, then by the same argument as before V_1 has k-r+1 distinct roots and therefore V has k-r+2 distinct latent roots $\delta_0^*, \delta_1^*, \ldots, \delta_{k-r+1}^*$.

We shall not present all the details, but suffice it to say that the form of Q in (III.B.3) would be duplicated in the present case with the sole difference that the first δ_n^* term involves all the correlation effects, all of which are non-estimable. Also the only terms that do enter implicitly into $\delta_1^*, \ldots, \delta_{k-r+1}^*$ terms and for which estimates can be found are the (a_i-b_i)'s, $(b=r+1,\ldots,k+1), b_{k+1}= 0$.

The constant correlations do not affect the result that least squares estimators for estimable functions of γ_i's are b.l.u. estimators of the same estimable functions. Using the arguments of the previous section we can prove

Theorem 19: Consider the class of mixed model situations described by (III.C.2) with the additional restrictions (III.C.5), (III.C.6) and $P_iX_j = 0$ $(i\neq 0) \neq (j\neq k+1)$. If we assume multivariate normality of β_i vectors as well, then the set of estimators $(X\hat{\gamma}_{LS}, s^2_{r+1},\ldots,s^2_{k+1})$ are complete sufficient for the parameters $(X\gamma, a_{r+1}-b_{r+1},\ldots,a_k-b_k,a_{k+1})$. Alternatively: If we consider the class of mixed models represented by (III.C.2) with the additional restrictions (III.C.6), $VXX' = XX'V$, $P_jX = 0$ $(j=r+1,\ldots k)$,

$$X_iX_i'X_jX_j' = X_jX_j'X_iX_i' \quad (i,j = r+1,\ldots k+1)$$ and

we assume multivariate normality of β_i vectors, then the set of estimators $(X\hat{\gamma}_{LS}, s^2_{r+1},\ldots,s^2_{k+1})$ are complete sufficient for the parameters $(X\gamma, a_{r+1}-b_{r+1},\ldots,a_k-b_k,a_{k+1})$.

In other words the usual estimators are M.V. also for the situation above, and may be said to be robust to constant correlations of random β_i effects ($i = r+1,\ldots,k$).

We conclude this section with the counterpart of Theorem 16.

Theorem 20: Suppose that within the model (III.C.2) we consider that class for with complete sufficiency can be established under normality, and restrict ourselves to cases in which projection operators satisfy the restrictions discussed in theorem 18. If we remove normality, but retain independence of vectors β_i and β_j for $(i \neq j)$ and if furthermore fourth moments are finite for all elements within a given vector β_i and if $E(\beta_{ik}^4) = c_i$, $E(\beta_{ik}^3) = d_i$, $E(\beta_{ik}^3 \beta_{il}) = e_i$, $E\, \beta_{ik}^2 \beta_{il} \beta_{im}) = f_i$, and $E(\beta_{ik} \beta_{il} \beta_{im} \beta_{in}) = g_i$ where $i = r+1,\ldots,k$, $c_{k+1} = f_{r+1} = g_{k+1} = 0$,

k, l, m and n are all unequal

and

$e_i, e_i, \ldots g_i$ are constants,

then the same estimators that are M.V. under normality are b.q.u. and b.l.u. for "variance components" and estimable functions of regression parameters respectively.

IV. ON ESTIMATION IN DESIGNED VARIANCE COMPONENTS MODELS

A. Introduction

In this chapter we consider situations which are known as "balanced" in the literature and which are designed in some special way, for example, b.i.b.'s and p.b.i.b.'s. The b.i.b.'s, for instance, have a relatively simple yet approximate analysis under the assumption that (say) treatments are fixed and blocks are random. Our interest in this chapter is to consider estimation in such "balanced designed" cases under a variance components model. The cases that interest us here are those that cannot be treated by the methods of Chapter III, i.e., those cases where for some i,j, the commutativity condition does not hold. In view of Theorem 3 , we recognize that simultaneous diagonalization independently of the parameters of the covariance matrix V by an orthogonal P is not possible and we have to have recourse to other methods.

The question that arises is whether U.M.V. estimators exist. We have no proof that they do not. However the minimal sufficient sets of statistics that have been exhibited thus far are known not to be complete. For competitive estimators, such as those exhibited by Bush and Anderson (1963), in different regions of the parameter space, different estimators have smallest variance. Furthermore we note that for finite samples maximum likelihood (M.L.) estimators do not have the good properties that they have when sample size is infinite, see Cramer (1946) and Lehmann (1950) for examples where M.L. estimators are inefficient for finite samples. There does in fact appear to be no alternative but to attempt to find the minimum variance estimator for every special situation that arises by trial and error. Obviously an undertaking of this nature depends rather crucially on having available formulae by which to evaluate the variances of estimators of variance components that are given by different methods. In Chapter V we contribute to the problem of finding the variance of a fairly general quadratic form. In the present chapter we attempt to clarify some aspects of a method of estimation that has long been known, and is commonly referred to as "least squares estimation method." We believe that the scope of this method is not fully appreciated and utilized. We point to the fact that Henderson (1953) could claim that the variances of estimators obtained by this method (his Method 3) were unknown. There have been some attempts for e.g. Searle (1956, 1958, 1961), Manamunulu (1963), at variances of variance components for estimators obtained by Henderson's (1953) method 1, and there has been an attempt at the variances of method 3 estimators of variance components in a 2-way unbalanced classification by Bush and Anderson (1963) who use a theoretical argument due to Roy (1957). In this chapter we present a "transformation method" for obtaining estimators through the use of independent single degree of freedom sums of squares. The novel feature of the method is that it enables us to find variances of variance components obtained by least squares quite easily; we think this alone makes the "method" worthy of presentation. We have decided to call this "method" the sequential transformation method.

B. On Estimation in the B.I.B. with Treatments Fixed

Rao (1947a) in discussing the b.i.b. design with treatments fixed, remarked that there are "best" estimators although the equations leading to them are complicated. If we are correct in presuming that Rao (loc. cit.) had M.L. estimators in mind, then the assertion appears to have no solid basis. In fact the only claim that can validly be made for M.L. estimators is that they are based on a minimal sufficient set.

Graybill and Weeks (1959) show that Yates' (1940) combined estimator for treatment contrasts which Rao (1947a) endorsed as a good approximate method is also based on a minimal sufficient set. The least squares estimators for σ_b^2 and σ^2 are also based on the minimal sufficient set of statistics. Whereas σ_B^2 is a function that involves all members of the set, σ^2 is a function of all members only if we allow some coefficients to be zero. As is well known, the question of exactly how a 'good' estimator should be constructed from a minimal sufficient set is an unsolved one.

In the b.i.b. with blocks and treatments random the same difficulty arises. Weeks and Graybill (1961) have exhibited a minimal sufficient set $(s_1,\ldots,s_6)$ for this case, but not shown how to use them to form "good" estimators. Estimators obtained by "least squares" in this case are functions of the set they give if we say that s_4 enters all functions with coefficient zero.

C. Algebraic Restatement of Method 3 of Henderson for Finding Variance Component Estimators in Completely Random Models

We shall have need in this section of the concept of a conditional inverse. It has been indicated by Rao (1962b) and others in unpublished material that considerable unification within the theory of least squares is made possible by a concept of this type.

Definition. Let A be any matrix. The matrix A^* is said to be a conditional inverse of A if A^* satisfies the relation $AA^*A = A$.

The concept of a conditional inverse finds theoretical utility in the solution of linear equations. We illustrate this in

Lemma 5: If A^* is a conditional inverse of A, and if $Ax=y$ are consistent then the vector A^*y is a solution to these equations.

Proof: Since the equations are consistent there exists a vector, x_0 say, such that

$$y = Ax_0 = AA^*Ax_0$$
$$= A(A^*y) ,$$

so that

A^*y is a solution to the equations

$Ax = y$.

A theoretical method for finding the conditional inverse of a n x p matrix A is to first find the non-singular matrices B and C such that

$$BAC = D$$

where D is the n x p matrix

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

i.e., the matrix with r^{th} order unit matrix in the upper left-hand position and zero elements everywhere else. It can be verified easily that the p x n matrix

$$D^* = \begin{bmatrix} I_r & U \\ W & V \end{bmatrix}$$

with the sub-matrices arbitrary, satisfies

$$DD^*D = D$$

and furthermore that the matrix $A^* = CD^*B$ is a conditional inverse for A. We know of no computer programs in existence for finding a conditional inverse in this way. It is interesting to note that the following well known procedure does in fact produce a conditional inverse for a symmetric matrix A (n x n). For convenience we consider A rearranged so that the independent rows and columns appear at the top left-hand corner. Delete dependent rows (and columns) of A until a matrix of full rank A_1 (say) is obtained. Then the n x n matrix

$$B = \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

is a conditional inverse of A. This follows because

$$ABA = \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix} \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix}$$

$$= \begin{bmatrix} A_1 & A_2 \\ A_2' & A_2'A_1^{-1}A_2 \end{bmatrix}$$

However, since rows deleted are linear combinations of rows remaining we can write

$(A_2' \; A_3) = P(A_1 \; A_2)$, i.e., $A_2' = PA_1$ and $A_3 = PA_2$. We have

$A_2' \, A_1^{-1}A_2 = PA_2 = A_3$ so that ABA = A and therefore B is a conditional inverse of A.

Another method for producing a conditional inverse for a symmetric matrix A is given by Rao (1962b).

We now consider the model

$$y = j_n\mu + \sum_{i=1}^{k+1} X_i \beta_i ,$$

$X_{k+1} = I$, $E(\beta_i \beta_j') = 0$ $(i \neq j)$, $E(\beta_i \beta_i') = I\,\sigma_i^2$

and for at least some $i \neq j$, $(i,j=0,\ldots,k+1)$,

$$X_i X_i' X_j X_j' \neq X_j X_j' X_i X_i' \quad \text{(IV.C.1)} .$$

We remark that a μ-term in (IV.C.1) is not essential to the arguments that follow.

In the model (IV.C.1), the variance

$$V = \{v_{ij}\} = \sum_{i=1}^{k+1} X_i X_i' \sigma_i^2$$

or for that matter, W, is not simultaneously diagonalizable independently of the parameters by an orthogonal matrix P.

The character of the general "approximate" method of analysis for models of the above type is to regard all effects as fixed, unknown constants, fit constants by least squares (L.S.) and then to equate the S.S. obtained to their true expectations in terms of the actual model. Further specifications along with the above outline were popularized by Henderson (1953) as Method 3, but the essentials of the method were mentioned in the literature by Wald (1947) and David and Johnston (1951, 1952) and were probably used long before 1947 as well. The details of Method 3 differ slightly from those of Kempthorne (1952), page 112. At this time our interest centers on the Method 3 of Henderson (1953) which we shall also refer to as the least squares (L.S.) method. No comprehensive analysis of the properties of the method, other than a demonstration that it is unbiased, appears to have been undertaken. One general aim of the present chapter is to contribute to such an analysis.

If in (IV.C.1) all β_i's except for β_{k+1} are regarded as fixed, and we fit all constants, then the L.S. estimator of the parametric vector fitted is given by any solution $\hat{\beta}$ to the equations

$$\begin{bmatrix} X_0'X_0 & X_0'X_1 \cdots X_0'X_k \\ X_1'X_0 & X_1'X_1 \cdots X_1'X_k \\ \cdot & \cdot \cdots \cdot \\ X_k'X_0 & X_k'X_1 \cdots X_k'X_k \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \cdot \\ \beta_k \end{bmatrix} = \begin{bmatrix} X_0'y \\ X_1'y \\ \cdot \\ X_k'y \end{bmatrix}$$

$$\text{or } (X'X)\beta = X'y .$$

The S.S. due to fitting all constants is

$$\begin{aligned} R(\mu, \beta_1, \beta_2, \cdots, \beta_k) &= \hat{\beta}'X'y \\ &= ((X'X)^*X'y)'\ X'y \\ &= y'X(X'X)^*\ X'y \ , \end{aligned}$$

where A^* represents a conditional inverse of A.

If in (IV.C.1) we ignore β_k, regard the remaining β_i's except β_{k+1} as fixed and fit constants, then the L.S. estimator of the parametric vector fitted is given by any solution $\tilde{\beta}$ to the equations

$$\begin{bmatrix} X_0'X_0 & X_0'X_1 & \cdots & X_0'X_{k-1} \\ X_1'X_0 & X_1'X_1 & \cdots & X_1'X_{i-1} \\ \cdot & \cdot & \cdots & \cdot \\ X_{k-1}X_0 & X_{k-1}X_1 & \cdots & X_{k-1}'X_{k-1} \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \\ \beta_{k-1} \end{bmatrix} = \begin{bmatrix} X_0'y \\ X_1'y \\ \\ X_{k-1}'y \end{bmatrix}$$

$$\text{or } (\overline{X}'\overline{X})\beta = \overline{X}'y .$$

The S.S. due to fitting all constants except β_k is

$$\begin{aligned} R(\mu,\beta_1,\beta_2,\beta_{k-1}) &= \tilde{\beta}'\overline{X}'y \\ &= ((\overline{X}'\overline{X})^*\overline{X}'y)'\ \overline{X}'y \\ &= y'\overline{X}(\overline{X}'\overline{X})^*\overline{X}'y . \end{aligned}$$

Finally, then, the S.S. due to fitting β_k is

$$\mathrm{Rem}(\beta_k) = y'X(X'X)^*X'y - y'\overline{X}(\overline{X}'\overline{X})^*\overline{X}'y \quad .$$

Henderson's (1953) Method 3 requires that Rem (β_i) be obtained for all i (i=1,...,k) and obtains point estimators of variance components as a simultaneous solution to the equations (i=1,...,k)

$$E(\mathrm{Rem}\ (\beta_i)) = \mathrm{Rem}(\beta_i) .$$

We notice that for the real matrices $A = X(X'X)^*X'$ and $B = \overline{X}(\overline{X}'\overline{X})^*X'$ $A' = A$, $B' = B$, $AA = A$ and $BB = B$, i.e. A and B are symmetric idempotent matrices. It is known that such matrices are orthogonal projection operators. Then another description of $R(\mu, \beta_1,\ldots, \beta_k)$ is that it consists of the square of the projection of y, on the column space of $(X_0,X_1,\ldots,X_k)$, while $R(\mu, \beta_1,\ldots,\beta_{k-1})$ is the square of the projection of y on the column space of $(X_0X_1 \ldots X_{k-1})$. The difference between two projections of y on different column spaces, one of which is a subspace of the other is itself a projection of y and the square of that projection is

$$\mathrm{Rem}(\beta_k) = R(\mu, \ldots, \beta_k) - R(\mu, \ldots, \beta_{k-1}) .$$

We shall make use of projection arguments to simplify expressions.

We are interested in (a) the expectation under a completely random model of the form (IV.C.1) of quadratic forms, i.e., $\mathrm{Rem}(\beta_i)$'s or lengths of projections, and (b) the variance of these quadratic forms when normality is assumed and when it is not. First we restate the solution to (a).

We have $E(\mathrm{Rem}(\beta_k)) = E(\hat{\beta}'X'y) - E(\bar{\beta}'\bar{X}'y)$

$$= E(y'Ay) - E(y'By)$$

where $A = X'(X'X)^*X'$ and $X = (X_0X_1 \ldots X_k)$,

and $B = \bar{X}(\bar{X}'\bar{X})^*\bar{X}'$ and $\bar{X} = (X_0X_1 \ldots X_{k-1})$.

Now $E(y'AY = E(\sum_{ij} a_{ij}y_iy_j) = \sum_{ij} a_{ij}(v_{ij}+E(y_i)E(y_j))$

$$= tr(AV) + (X_0\mu)'X'(X'X)^*X'X_0\mu$$

$$= tr(AV) + (X_0\mu)'(X_0\mu)$$

$= tr(AV) + n\mu^2$, where $tr(A)$ = trace of matrix A.

Likewise

$$E(y'By) = tr(BV) + n\mu^2$$

so that

$$E(y'Ay-y'By) = tr(A-B)V .$$

Since the matrix V has the form $X_1X'_1\sigma^2_1 + X_2X'_2\sigma^2_2 + .. + I\sigma^2_{k+1}$

$$\begin{aligned}tr(A-B)V = {} & tr(X(X'X)^*X'X_1X_1' - \overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_1X_1')\,\sigma_1^2 \\ & + tr(X(X'X)^*X'X_2X_2' - \overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_2X_2')\,\sigma_2^2 \\ & + \ldots\ldots\ldots\ldots\ldots\ldots \\ & + tr(X(X'X)^*X' - \overline{X}(\overline{X}'\overline{X})^*\overline{X}')\,\sigma_{k+1}^2 \qquad (IV.C.2)\ .\end{aligned}$$

If the column space of X_i is a sub-space of both X and $\overline{X}$ then the expectation will not involve σ_i^2 since in that case

$$X(X'X)^*X'X_i = \overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_i = X_i$$

and the coefficient of σ_i^2 drops out of the above expression. When the column space of X_i is a subspace of X, but not of $\overline{X}$, then we have a partial simplification in the coefficient of σ_i^2 to

$$\begin{aligned}& tr(X_iX_i' - \overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_iX_i') \\ & = N - tr(\overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_iX_i')\ .\end{aligned}$$

We notice that in (IV.C.2) only σ_k^2 has a coefficient of the latter type. Because $X(X'X)^*X'$ and $\overline{X}(\overline{X}'\overline{X})^*\overline{X}'$ are symmetric idempotent matrices

$$tr(X(X'X)^*X') = rk(X(X'X)^*X') = rk(X) \text{ and}$$
$$tr(\overline{X}(\overline{X}'\overline{X})^*\overline{X}') = rk(\overline{X}(\overline{X}'\overline{X})^*\overline{X}') = rk(\overline{X})$$

so that the coefficient of σ_{k+1}^2 is always the difference in ranks of the matrices X and $\overline{X}$. Therefore

$$E(Rem(\beta_k)) = (N - tr(\overline{X}(\overline{X}'\overline{X})^*\overline{X}'X_kX_k')\sigma_k^2 - (rk(X) - rk(\overline{X}))\sigma_{k+1}^2$$

The residual S.S. is given by

$$R = y'y - y'X(X'X)^*X'y$$

so that

$$\begin{aligned} E(R) &= tr(X_1X_1' - X(X'X)^*X'X_1X_1')\,\sigma_1^2 + \ldots + tr(I-X(X'X)^*X')\sigma_{k+1}^2 \\ &= \qquad 0 + \quad . \quad . \quad . \quad + (n-p)\,\sigma_{k+1}^2 \\ &= (n-p)\,\sigma_{k+1}^2 \quad \text{where } rk(X) = p\,. \end{aligned}$$

In perhaps the majority of cases there will be substantial simplification, and the need for determining a conditional inverse does not arise. In those cases where no simplification is possible we may use the result stated at the beginning of this section. We point out that (IV.C.2) and subsequent simplifications is merely a reformulation of Henderson's (1953) equation (19). Another form is given by Bush and Anderson (1963). One of the merits of the above form (shared by some others also) is that it is amenable to computations on an electronic machine.

We conclude this section with a summary of available results to (b). If we are prepared to assume normality of distribution of the y vector then the variances of the estimator given by

$$Rem(\beta_k) = y'(A-B)y = (y-\mu)'(A-B)(y-\mu) \quad \text{is}$$

$$\begin{aligned} &Var((y-\mu)'(A-B)(y-\mu)) \\ &\quad = 2\,tr(AV)^2 + 2\,tr(BV)^2 - 4\,tr(AVBV)\,. \end{aligned}$$

This result is due to Matern (1949). A more general formula that does not assume normality is derived in Chapter V.

The results of Chapter V allow us to obtain also

$$\mathrm{Cov}(y'(A-B)y,\ y'(A-c)y) = w_{12} \text{ (say)}$$

so that the covariance matrix of a succession of linear functions of estimators of the variance components $U\alpha$ (say) where $\alpha = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{k+1}^2)'$ is available. Then

$$\mathrm{Var}(\alpha) = U^{-1}\Omega U^{-1'} \text{ where } \Omega = \{w_{ij}\} .$$

We could measure the overall effectiveness of Method 3 of Henderson (1953), and any competitor, by some function of Var (α) such as

$$\mathrm{tr}(U^{-1}\Omega U^{-1'}) \text{ or } \det(U^{-1}\Omega U^{-1'}) .$$

We favor these two measures because they are sensitive to gross imprecision in estimation of individual components. From some points of view, a method that is found to give reasonable precision for all estimators is a desirable one. Bush and Anderson (1963) present some empirical evidence supporting the contention that Method 3 is such an estimator.

D. The Transformation Approach to Estimation in a Special Case

1. Deriving estimators for the b.i.b. with random blocks and treatments

A balanced incomplete block is defined as a design in which there are t treatments and b blocks of k experimental units per block (where $k < t$) with each treatment replicated r times. The arrangement of treatments is such that every pair of treatments occurs together in exactly λ blocks and each treatment occurs in a block only once. The model may be written as a special case of the general two-way classification given by Kempthorne (1952) for example, i.e.

$$y_{ijm} = \mu + \alpha_i + \tau_j + e_{ijm} \qquad \text{(IV.D.1)}$$

where $i = 1,\ldots,b$; $j = 1,2,\ldots,t$; $m = n_{ij}$;

where

$$n_{ij} = \begin{cases} 1 & \text{if treatment } j \text{ occurs in block } i \\ 0 & \text{otherwise.} \end{cases}$$

It should be noted that only the y_{ijm} in which $m \neq 0$ are observed.

Equation (IV.D.1) represents $bk = n$ equations, and these equations may be written in matrix form as

$$y = X_0 u + X_1 \beta_1 + X_2 \beta_2 + X_3 \beta_3$$

where the dimensions of the matrices are: $y(n \times 1)$; $X_1(n \times b)$; $\beta_1(b \times 1)$; $X_2(n \times t)$; $\beta_2(t \times 1)$; $\beta_3(n \times 1)$, and $X_0 = j_n$, $X_3 = I(n \times n)$. Furthermore, we notice that for at least some $i \neq j$

$$X_i X_i' X_j X_j' \neq X_j X_j' X_i X_i' \quad (i,j = 0,1,2,3) \qquad \text{(IV.D.2)} .$$

Although we can write down an orthogonal P $(n \times n)$ and consequently a "single degree of freedom breakdown"

$$y'y = y'P'Py = \sum_{i=1}^{n} y'P_i'P_i y$$

in view of lack of commutativity it does not necessarily follow that all the terms $y'P_i'P_i y$ will then have zero covariances and it is not possible to identify subspaces of P that correspond to similar latent roots of V since the latent roots are not obtainable without specifying the unknown parameters.

Henderson's (1953) method 3 or the least squares method applied as though all effects are constant, actually selects projection operators orthogonal to collections of column spaces of X_i matrices. Although this may therefore give a method of collecting subspaces of P to form lines of an A.o.V. table, it is apparent that the expectations of single degrees of freedom that are gathered together to form lines of an A.o.V. table are not necessarily homogenous. There are reasons for wanting to know the expectations of constituent parts of a line. The following sequential transformation method exhibits a way of actually determining a single degree of freedom breakdown and the expectations of each quadratic form ($y'P_i'P_iy$). It also allows us to determine variances of estimators quite easily.

The details immediately following are not essential, but are useful for computation purposes. In specifying the order of observations in the vector y of (IV.D.1) we let this be determined by the parameter β_1. Thus, for now, we require that all observations that incur the first value of the β_1 effect occur prior to all those that incur some other

β_1 effect, and so on. In this way we ensure that the structure of X_1X_1' is a diagonal block of b J_k^k matrices. We shall use the notation y^b to denote that the β_1 (or the b parameter) has determined the order in which observations enter y. It is known that there exists an orthogonal matrix P such that P $X_1X_1'P'$ is diagonal and that there are only two distinct latent roots; b roots will in fact equal k and n-b will equal zero. Let P_1 denote the matrix of vectors that correspond to the root k. The matrix formed from the remaining vectors of P is denoted by P_2.

The two matrices P_1 and P_2 suggest a transformation of the vector y, into two equivalent sets P_1y^b and P_2y^b. Now the covariance matrix of P_1y^b is

$$V_1 = k\ I_b\ \sigma^2_{\beta_1} + P_1X_2X_2'P_1'\ \sigma^2_{\beta_2} + I_b\ \sigma^2_{\beta_3}$$

and the covariance matrix of P_2y^b is

$$V_2 = \qquad P_2X_2X_2'P_2'\ \sigma^2_{\beta_2} + I_{n-b}\ \sigma^2_{\beta_3}\ .$$

Analysis then proceeds separately with the sets P_1y^b and P_2y^b as though each of these constitutes all the data. Consider first V_1. There exists an orthogonal matrix Q such that QV_1Q' is diagonal. In fact we would determine Q in practice as the orthogonal matrix that diagonalizes $P_1X_2X_2'P_1'$. This means that if we transform P_1y^b to QP_1y then this latter set of b orthogonal linear forms has diagonal covariance matrix QV_1Q'. Next we bring the transforms P_2y^b to a similar stage, and we then indicate how we propose to get estimators.

There exists an orthogonal matrix R such that RV_2R' is diagonal. This means that if we transform P_2y to RP_2y^b then this set of (n-b) orthogonal linear forms has diagonal covariance matrix RV_2R'.

We may verify that the matrix $A = \begin{bmatrix} QP_1 \\ \hline RP_2 \end{bmatrix}$ is

$b + n-b \ (=n) \times n$ and orthogonal; consequently if we obtain this matrix in the way described, then we have a breakdown $y'y = y'A'Ay = \sum_{i=1}^{n} y'A_i'A_iy$ where A_i $(i=1,\ldots,n)$ is the i^{th} row of A. Furthermore, since we have the individual $\mathrm{Var}\,(A_iy^b)$ $(i=1,\ldots,n)$ terms, it is apparent which rows of A_i can be used simultaneously for estimation purposes. This collecting of rows of A corresponding to similar expectations, gives rise to B_i $(i=1,\ldots,m)$ (say); then we have

$$y^{b'}y^b = \sum_{i=1}^{m} y^{b'}B_i'B_iy^b$$

where $E(y^{b'}B_i'B_iy^b)$ is some homogenous linear function of variance components, and only $y^{b'}B_1'B_1y^b$ involves μ^2 in its expectation. We propose to obtain estimators by equating some quadratic forms of type $y^{b'}B_i'B_iy^b$ to their expectations and solving the resulting equations. Various different sets of estimators for $\sigma^2_{\beta_1}$, $\sigma^2_{\beta_2}$ and $\sigma^2_{\beta_3}$ can be envisaged in this way.

It is clear from an investigation of the matrices V_1 and V_2 above that an estimate of $\sigma^2_{\beta_2}$ and $\sigma^2_{\beta_3}$ can be obtained from the transformed data $R_2P_2y^b$, while the estimate of $\sigma^2_{\beta_1}$ from QP_1y cannot in general be similarly disentangled of $\sigma^2_{\beta_2}$ terms. We point out that in order to estimate $\sigma^2_{\beta_1}$ in the way of $\sigma^2_{\beta_2}$ above we would repeat the procedure (followed to obtain Ay) after insisting on an appropriate occurrence of components of the β_2 vector in the original equations by setting up y in the appropriate way. We distinguish the order of observations in y with β_2 in mind by y^t.

We have described the method for a model of type (IV. D.2). The method can be extended without difficulty to models that involve more than three variance components. In the previous section we described how to obtain $\text{Rem}(\beta_k)$ in the model (IV.C.2). The transformation method appropriately applied in that case would in fact isolate the single degree of freedom sums of squares that make up $\text{Rem}(\beta_k)$. This would have been achieved by choosing P such that $P(X_0 \ldots X_{k-1})(X_0 \ldots X_{k-1})'P' = \Lambda$ diagonal and denoting those vectors of P that correspond to zero latent roots by P_2.

2. A special example

We give an abbreviated version of the detailed analysis by the method of sub-section 1 in the case of a b.i.b.d. with parameters $b = 6$, $t = 4$, $k = 2$, $r = 3$ and $\lambda = 1$.

We obtain quite easily in succession a matrix $P = (P_1{}' \vdots P_2{}')'$, a matrix $P_1X_2X_2{}'P_1{}'$ and a matrix Q such that $QP_1X_2X_2{}'P_1Q'$ is diagonal. Likewise the form of $P_2X_2X_2{}'P_2{}'$ is easily available, and a matrix R such that $RP_2X_2X_2{}'P_2{}'R' = \Lambda$. We obtain for $Ay^b = \begin{bmatrix} QP_1 \\ RP_2 \end{bmatrix} y^b$ the form

$$\left[\begin{array}{cccccccccccc}
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\
\hline
\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
\hline
-\frac{2}{2\sqrt{6}} & -\frac{2}{2\sqrt{6}} & -\frac{2}{2\sqrt{6}} & -\frac{2}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} \\
\cdot & \cdot & \cdot & \cdot & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\
\hline
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & \cdot & \cdot & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\
\hline
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \cdot & \cdot & \cdot \\
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \cdot & \cdot & \cdot & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\
\cdot & \cdot & \cdot & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}}
\end{array}\right] y^b = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \end{bmatrix} y^b$$

with covariance matrices of QP_1y^b equal to

$$\begin{bmatrix} 2\sigma_B^2+3\sigma_T^2+\sigma^2 & & & & & \\ & 2\sigma_B^2+\sigma_T^2+\sigma^2 & & & & \\ & & 2\sigma_B^2+\sigma_T^2+\sigma^2 & & & \\ & & & 2\sigma_B^2+\sigma_T^2+\sigma^2 & & \\ & & & & 2\sigma_B^2+\sigma^2 & \\ & & & & & 2\sigma_B^2+\sigma^2 \end{bmatrix}$$

and of RP_2y^b equal to

$$\begin{bmatrix} 2\sigma_T^2+\sigma^2 & & & & & \\ & 2\sigma_T^2+\sigma^2 & & & & \\ & & 2\sigma_T^2+\sigma^2 & & & \\ & & & \sigma^2 & & \\ & & & & \sigma^2 & \\ & & & & & \sigma^2 \end{bmatrix}$$

where we have of course assumed that y^b is written out appropriately for β_1 and have put $\sigma_{\beta_1}^2 = \sigma_B^2$, $\sigma_{\beta_2}^2 = \sigma_T^2$ and $\sigma_{\beta_3}^2 = \sigma^2$.

If we order the vector y so that all those observations that incur the first component of the vector β_2 occur before those that incur the second component of β_2 and so on we derive a different orthogonal (A) matrix, which we shall denote by $\bar{A}$.

In this case, $\bar{P}$ is a matrix that diagonalizes X_2X_2', where $\bar{P}_1$ corresponds to the non-zero latent root and $\bar{P}_2$ corresponds to the zero latent root.

The covariance matrix of $\bar{P}_1y^t$ is

$$\bar{V}_1 = \bar{P}_1X_1X_1'\bar{P}_1' \, \sigma^2_{\beta_1} + rI_t \, \sigma^2_{\beta_2} + I_t \, \sigma^2_{\beta_3}$$

and the covariance matrix of $\bar{P}_2y^t$ is

$$V_2 = \bar{P}_2 \, X_1X_1'\bar{P}_2' \, \sigma^2_{\beta_1} + I_{n-t} \, \sigma^2_{\beta_3} .$$

In this case we obtain in order the matrix $\bar{P}$, the matrix $\bar{P}_1X_1X_1'\bar{P}_1$ and a matrix $\bar{Q}$ such that

$$\bar{Q}\bar{P}_1X_1X_1'\bar{P}_1\bar{Q}' \quad \text{is diagonal.}$$

We then obtain $\bar{P}_2X_1X_1'\bar{P}_2'$ and a matrix $\bar{R}$ such that

$$\bar{R} \, \bar{P}_2X_1X_1'P_2\bar{R}' \text{ is diagonal.}$$

In the example above we obtain

$$\bar{A}y^t = \begin{bmatrix} \bar{Q} \, \bar{P}_1 \\ \bar{R} \, \bar{P}_2 \end{bmatrix} y^t =$$

$$\left[\begin{array}{cccccccccccc}
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\
\hline
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \\
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\
-\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \\
\hline
\cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot \\
\frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \cdot & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} \\
\hline
\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdot \\
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} \\
\hline
\frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{2}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{12}} \\
-\frac{8}{4\sqrt{6}} & \frac{10}{4\sqrt{6}} & -\frac{2}{4\sqrt{6}} & \frac{8}{4\sqrt{6}} & -\frac{10}{4\sqrt{6}} & \frac{2}{4\sqrt{6}} & -\frac{8}{4\sqrt{6}} & \frac{10}{4\sqrt{6}} & -\frac{2}{4\sqrt{6}} & \frac{8}{4\sqrt{6}} & -\frac{10}{4\sqrt{6}} & \frac{2}{4\sqrt{6}} \\
\frac{10}{4\sqrt{6}} & -\frac{8}{4\sqrt{6}} & \frac{2}{4\sqrt{6}} & \frac{10}{4\sqrt{6}} & -\frac{8}{4\sqrt{6}} & \frac{2}{4\sqrt{6}} & \frac{10}{4\sqrt{6}} & -\frac{8}{4\sqrt{6}} & \frac{2}{4\sqrt{6}} & \frac{10}{4\sqrt{6}} & \frac{8}{4\sqrt{6}} & \frac{2}{4\sqrt{6}}
\end{array}\right] y^t = \begin{bmatrix} \overline{B}_1 \\ \overline{B}_2 \\ \overline{B}_5 \\ \overline{B}_3 \\ \overline{B}_4 \end{bmatrix} y^t$$

with covariance matrix of $\overline{Q}\overline{P}_1 y^t$ equal to

$$\begin{bmatrix} 2\sigma_B^2+2\sigma_T^2+\sigma^2 & & & \\ & 2/3\sigma_B^2+3\sigma_T^2+\sigma^2 & & \\ & & 2/3\sigma_B^2+3\sigma_T^2+\sigma^2 & \\ & & & 2/3\sigma_B^2+3\sigma_T^2+\sigma^2 \end{bmatrix}$$

and of $\overline{R}\overline{P}_2 y^t$ equal to

$$\begin{bmatrix} \sigma^2 & & & & & & & \\ & \sigma^2 & & & & & & \\ & & \sigma^2 & & & & & \\ & & & 2\sigma_B^2+\sigma^2 & & & & \\ & & & & 2\sigma_B^2+\sigma^2 & & & \\ & & & & & 4/3\sigma_B^2+\sigma^2 & & \\ & & & & & & 4/3\sigma_B^2+\sigma^2 & \\ & & & & & & & 4/3\sigma_B^2+\sigma^2 \end{bmatrix}.$$

Although there are alternatives, and we shall mention at least one (Method 1) we favor estimating σ_B^2 from $\overline{R}\overline{P}_y^{-t}$. Having decided to do this, then further alternatives arise. Firstly we may decide on straightforward pooling of the single degree of freedom sums of squares i.e.

$$y^{t\prime}\overline{B}_3{}'\overline{B}_3 y^t + y^{t\prime}\overline{B}_4{}'\overline{B}_4 y^t$$

and subtract an appropriate multiple of $y^{t\prime}\overline{B}_5{}'\overline{B}_5 y^t$ to give an estimate of $c\sigma_B^2$. This is the conventional estimator; we notice that under an assumption of normality the quadratic forms $y^{t\prime}\overline{B}_5{}'\overline{B}_5{}' y^t$, $y^{t\prime}\overline{B}_3{}'\overline{B}_3 y^t$ and $y^{t\prime}\overline{B}_4{}'\overline{B}_4 y^t$ have χ^2

distributions. We note that under the assumption of a known covariance structure for the relevant mean squares a weighted estimator can be found which in general will have smaller variance than the one above.

We may make use of generalized least squares to obtain a b.l.u. estimator of σ_B^2 as follows. Set up the model linear in the parameters σ^2 and σ_B^2

$$M = (M_1' M_2' M_3')' = Z(\sigma^2, \sigma_B^2)' + e$$

where $M_i = (\overline{R}_i \overline{P}_2 y^t)' (\overline{R}_i \overline{P}_2 y^t)$ $i=1,2,3,$ $E(e) = 0$, and $E(ee') = D$ (diagonal) $= Var(M)$, and Z is a known matrix. With D known the b.l.u. estimator of $\gamma = (\sigma^2 \sigma_B^2)'$ is given by

$$(Z'D^{-1}Z)^{-1} Z'D^{-1}M \quad \text{with variance } (Z'D^{-1}Z)^{-1} .$$

When D is unknown, we would use some initial estimator for D, and proceed by an iterative process to a reasonable solution for σ_B^2 .

One of the basic unsolved problems in the estimation of variance components, is the loss involved in confining ourselves to M in order to estimate σ_B^2. Another difficulty arises in the case where we use estimated weights to form weighted estimators. Some discussion of the inaccuracy incurred in doing this is given by Kempthorne (1952) page 463.

We notice that both analyses give a breakdown into five associations of vectors (i.e. B_i's and $\overline{B}_i$'s) of the orthogonal matrices A and $\bar{A}$. Thus for Ay^b we wrote $(B_1'B_2'B_3'B_4'B_5')'y^b$ where for example B_1 refers to the first vector of A while B_4 refers to the seventh, eighth and ninth vectors of A. We gave the same subscripts to $\bar{B}$ vectors, where possible, as in B. Thus $\bar{B}_1$ corresponds to the "mean" and $\bar{B}_5$ to "error" as before.

Some possible estimators that utilize simple weighting are:

$\hat{\mu}$: $\bar{y}$

$\hat{\sigma}^2$: $y^{b'}B_5'B_5y^b/3$ or $y^{t'}\bar{B}_5'\bar{B}_5y^t/3$

$\hat{\sigma}_T^2$: $y^{b'}B_4'B_4y^b - 3\hat{\sigma}^2/6$; $(y^{t'}\bar{B}_2'\bar{B}_2y^t - 2\hat{\sigma}_B^2 - 3\hat{\sigma}^2)/9$

$\hat{\sigma}_B^2$: $(y^{b'}B_2'B_2y^b + y^{b'}B_3'B_3y^b - 3\hat{\sigma}_T^2 - 5\hat{\sigma}^2)/10$;

$(y^{t'}\bar{B}_3'B_3y^t + y^{t'}\bar{B}_4'\bar{B}_4y^t - 5\hat{\sigma}^2)/8$.

The only way in which we can decide between these estimators (and others) is by means of the variances of the estimators. We return to this later.

In the example above, we found that the expectation of each of the (t-1) sums of squares that go to make up the blocks eliminating treatments component were homogeneous. The coefficient of σ_B^2 was in agreement with the well-known "average" value for this coefficient, namely $Ek = (k-1)t/(t-1)$ from a b.i.b. with treatments fixed.

We may verify that in the general b.i.b. design the single degree of freedom component sums of squares of the treatment component of blocks eliminating treatments are homogeneous in their expectation, by noting that each such degree of freedom is a linear function of contrasts of intrablock and interblock estimators of treatment effects. It can be verified that the expectation of any such single degree of freedom sum of squares is $\sigma^2 + Ek\ \sigma_b^2$.

An alternative method by which verification could take place for any particular b.i.b. design is given below. Each matrix entering $\overline{P}_2 X_1 X_1' \overline{P}_2'$ is real, so that the product is real. The product is also symmetric and therefore diagonalizable.

Theorem 21 (Mirsky (1961)): A is similar to a diagonal matrix if and only if rank $(\alpha I - A) = n - m_\alpha(A)$ for every value α, where $m_\alpha(A)$ is the multiplicity of α as a characteristic root of A.

Write $\overline{P}_2 X_1 X_1' \overline{P}_2' = A$, $Ek = \alpha$. Obtain $rk(\alpha I - A)$ for the particular situation on hand. If rank $(\alpha I - A) = b - t$ we conclude that the (t-1) degrees of freedom in question are homogeneous in their expectation. It should be noted that the above two methods are generally applicable to the verification of homogeneity of expectations of arbitrary single degrees of freedom sums of squares in any given mean square.

3. Comparison of methods

Henderson (1953) suggested two sets of estimators for models of the type (IV.D.1). Method one suggested the use of sums of squares for blocks (treatments) ignoring treatments (blocks). Method 3 suggested fitting constants or L.S. i.e. proceeding as though the parameters were fixed and equating sums of squares to expectations.

It is clear that a combination of methods 1 and 3 may be regarded as an orthogonal breakdown of the total sum of squares for a model of type (IV.C.2) into

$$y'y = y'C_1y + y'C_2y$$

where C_1 is a projection on the column space of the matrix $(X_0X_1X_2\ldots X_{k-1})$ and C_2 is a projection on the column subspace of $(X_0X_1\ldots X_k)$ which is orthogonal to the space $(X_0X_1\ldots X_{k-1})$.

An examination of the matrices A and $\overline{A}$ derived by the transformation method reveals that they are in fact constructed to achieve an orthogonal breakdown of this type. We may verify that the breakdowns suggested by the transformation method and those of Henderson (1953), in a b.i.b., are in agreement by showing that

E (blocks ignoring treatments)= $E(y^{b'}B_2'B_2y^b+y^{b'}B_3'B_3y^b)$

E (treatments ignoring blocks) = $E(y^{t'}\overline{B}_2'\overline{B}_2y^t)$

E (blocks eliminating treatments) = $E(y^{t'}\overline{B}_3'\overline{B}_3y^t+y^{t'}\overline{B}_4'\overline{B}_4y^t)$

E (treatments eliminating blocks) = $E(y^{b'}B_4'B_4y^b)$.

Weeks and Graybill (1961) give a minimal set of sufficient statistics for the b.i.b. design, but give no estimators. The interesting thing about the set which they derive is that the expectation of s_5 agrees exactly with the expectation of treatments eliminating blocks which the Method 3 of Henderson (1953) would suggest to estimate σ_T^2. Of course, in the latter method the concept of degrees of freedom is well defined whereas in the minimal set as defined by Weeks and Graybill (1961) the concept of degrees of freedom is not mentioned.

If we allow symmetry to influence our choice of an estimator for σ_B^2, then we arrive at the Henderson (1953) Method 3 estimator for this component as well. If we give zero degrees of freedom to s_4 and (t-1) to s_3 then the estimator that the minimal sufficient set would appear to suggest for σ_B^2 is not the Henderson Method 3 estimator but the one given by Method 1.

4. Variances of estimators

An advantage of the transformation method arises from having available the matrices of the transformation i.e. the B_i's. This allows us to obtain quite simply the variances of Method 3 estimators of the variable components. For example in sub-section 3 we suggested the estimator

$$\hat{\sigma}_T^2 = y^{b'}B_4'B_4y^b - y^{b'}B_5'B_5y^b/6 .$$

Hence, under normality

$$\mathrm{Var}(\hat{\sigma}_T^2) = (2\,\mathrm{tr}(VB_4'B_4)^2 + 2\,\mathrm{tr}(VB_5'B_5)^2)/36$$

where $V = X_1X_1'\sigma_B^2 + X_2X_2'\sigma_T^2 + I\sigma^2$. No covariance term does in fact enter on the right-hand side, because the linear functions B_4y and B_5y are chosen by the transformation method to be uncorrelated.

We note that an estimate of $\mathrm{Var}(\sigma_T^2)$ may be obtained by substituting estimates for variance components in the above expression for V.

V. ON ESTIMATION IN MIXED AND RANDOM UNBALANCED MODELS

A. On Least Squares Type Estimators of Variance Components and Variances of Estimators

1. Introduction

In this chapter we consider mixed variance component models that do not exhibit features under which the models would be either balanced$_2$ or "designed unbalanced". We have in mind, for example, random models to which concomitant parts have been added, and random models that have missing observations. There do not appear to have been any attempts to derive minimal sufficient sets of statistics for such models. It appears, however, that the condensation of information obtained in this way would not be very great and in any case the use of the minimal set to obtain "good" estimators presents an open problem.

We devote ourselves almost entirely to a method of estimation for these cases which is analogous to Henderson's (1953) Method 3, discussed in Chapter IV. This method will be seen to have some desirable features of simplification that will lead to formulae for variances of estimators of variance components.

In a mixed model, Method 3 of Henderson (1953) consists of fitting all constants as though the model were fixed and then fitting all constants but one. The difference in the sums of squares obtained is equated to the difference in expectation of the quadratic forms in question. This gives an equation that corresponds to the variance component of the factor ignored for the second fitting. When this procedure is repeated for every random factor an equation for every variance component is obtained. The set of equations is then solved simultaneously (we assume this to be possible i.e., all σ_i^2 are estimable) to give point estimators of all the variance components.

Henderson (1953) did not give any formulae for the variances of estimates obtained by the method, and none have since been given. One aim of the present chapter is to obtain formulae for variances and covariances of quadratic forms which arise in the execution of Henderson's (1953) Method 3 and in some other related methods.

The model considered is the familiar one, namely

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=r+1}^{k+1} X_i \beta_i$$

where $E(\beta_{k+1}\beta'_{k+1}) = I\,\sigma_{k+1}^2$ and $X_{k+1} = I$ (V.A.1).

We shall find it convenient for our present purposes to represent (V.A.1) also by the model

$$y = X_{(1)}\gamma + X_{(2)}\beta + e$$

where $X_{(1)} = (X_0 X_1 \ldots X_r)$,

$X_{(2)} = (X_{r+1} \ldots X_k)$,

$$\gamma = (\gamma_0' \gamma_1' \cdots \gamma_r')' \text{ , and } \beta = (\beta'_{r+1} \cdots \beta'_k)' .$$

Let $X = (X_0 X_1 \ldots X_r X_{r+1} \ldots X_{k+1})$ and

$$\bar{X} = (X_0 X_1 \ldots X_r X_{r+2} \ldots X_{k+1}) .$$

Then we may formalize the Method 3 estimation procedure for σ^2_{r+1} by writing

$$R(\gamma_0 \cdots \gamma_r \, \beta_{r+1} \cdots \beta_{k+1}) = A = y'X(X'X)^*X'y = y'My \quad \text{and}$$

$$R(\gamma_0 \cdots \gamma_r \, \beta_{r+2} \cdots \beta_{k+1}) = B = y'\bar{X}(\bar{X}'\bar{X})^*\bar{X}'y = y'Ny$$

where $X(X'X)^*X' = M$ and $\bar{X}(\bar{X}'\bar{X})^*\bar{X}' = N$.

If $M - N = T = t_{ij}$ say, then

$$\begin{aligned} \text{Rem}(\beta_{r+1}) &= y'(M-N)y = y'Ty \\ &= (X_{(1)}\gamma + X_{(2)}\beta + e)'T(X_{(1)}\gamma + X_{(2)}\beta + e) \\ &= (X_{r+1}\beta_{r+1} + e)'T(X_{r+1}\beta_{r+1} + e) \end{aligned}$$

Therefore

$$E(\text{Rem}(\beta_{r+1})) = \text{tr } T(X_{r+1}X'_{r+1}\sigma^2_{r+1} + I\sigma^2_{k+1}) .$$

Suppose we equate $E(\text{Rem}(\beta_{r+i}))$ and the corresponding $\text{Rem}(\beta_{r+i})$ values in the way described; we obtain

$$\begin{bmatrix} E(\text{Rem}(\beta_{r+1})) \\ E(\text{Rem}(\beta_{r+2})) \\ \cdots \\ E(\text{Rem}(\beta_{k+1})) \end{bmatrix} = \begin{bmatrix} u_{11} & & u_{1k-r} \\ & u_{22} & \\ & \cdot & \cdot \\ & & u_{k-r,k-r} \end{bmatrix} \begin{bmatrix} \sigma^2_{r+1} \\ \sigma^2_{r+2} \\ \cdot \quad \cdot \\ \sigma^2_{k+1} \end{bmatrix} .$$

Then

$$\begin{bmatrix} \hat{\sigma}^2_{r+1} \\ \hat{\sigma}^2_{r+2} \\ \cdot\ \cdot \\ \hat{\sigma}^2_{k+1} \end{bmatrix} = \begin{bmatrix} u^{11} & & u^{1,k-r} \\ & u^{22} & \\ & \cdot & \\ & & u^{k-r,k-r} \end{bmatrix} \begin{bmatrix} \mathrm{Rem}(\beta_{r+1}) \\ \mathrm{Rem}(\beta_{r+2}) \\ \cdot\ \cdot \\ \mathrm{Rem}(\beta_{k+1}) \end{bmatrix}$$

We require for example a general formula for

$$\mathrm{Cov}(\mathrm{Rem}(\beta_{r+i}), \mathrm{Rem}(\beta_{r+j})) \quad (i,j = 1,\ldots k-r+1)$$

and it is one purpose of the present chapter to obtain such a formula. We shall then be in a position to obtain the covariance matrix of the vector of estimators $(\hat{\sigma}^2_{r+1},\ldots,\hat{\sigma}^2_{k+1})'$ and we shall be in a position to make comparisons between estimators given by Method 3 and those that may be supplied by another method. In the section B we shall consider one other method of estimation in a random model that has fixed concomitants added to it, and will also find variances of some possible estimators in that case.

2. Variances and covariances of quadratic forms in mixed and random models

a. Introduction The approach followed is related to the one of David and Johnston (1951, 1952). It is, however, more general in that it applies, for example, to random classification models and it does not require the classification matrices to be expressed in full-rank form.

Clearly when Method 3 of Henderson (1953) is applied to a mixed model, insofar as estimators of variance components are concerned it is possible to simplify all quadratic forms so that effectively they do not involve constant factors. Actually then, in practice, finding variances and covariances of mean squares given by this method is no different in principle from finding variances of mean squares in a random model. However, because there may be some intrinsic interest in the general formula for the variance of a quadratic form arising in a mixed model we devote a section to obtaining such a form.

In sub-sub-section (b), to fix ideas, we consider finding the variance for a quadratic form in the case of a model involving a single random factor. In sub-sub-section (d) we consider a more general case of a quadratic form arising by Method 3 of Henderson (1953) in a model where β_j (say) represents an interaction of two earlier occurring β_i's. We arbitrarily restrict ourselves to classification models since these appear to be the primary area for application of these results. No difficulty is foreseen in extending the results to more general (regression) models.

b. <u>Single random factor case</u> Let the model be $y = X_{(1)}\gamma + X_{(2)}\beta + e$, where β denotes a single random factor. The residual S.S. assuming all factors fixed is

$$Sa = y'(I-M)y = (X_{(1)}\gamma+X_{(2)}\beta+e)'(I-M)(X_{(1)}\gamma+X_{(2)}\beta+e)$$

where $M = X(X'X)^*X'$, and $X = X_1X_2$. Since $(I-M)X = 0$, we have

$$S_a = (X'_{(2)}\beta+e)'(I-M)(X_{(2)}\beta+e) = e'(I-M)e \quad \text{and}$$

$$E(S_a) = tr(I-M)\ \sigma_e^2 \ .$$

The minimum $S.S._{H_1}$ or the residual S.S. for a model that fits only the fixed factors is

$$S_f = (X_{(1)}\gamma+X_{(2)}\beta+e)'(I-X_{(1)}(X'_{(1)}X_{(1)})^*X'_{(1)})(X_{(1)}\gamma+X_{(2)}\beta+e)$$

Write $X_{(1)}(X'_{(1)}X_{(1)})^*X'_{(1)} = N$, then

$$S_f = (X_{(2)}\beta+e)'(I-N)(X_{(2)}\beta+e) \ .$$

The S.S. due to β effects, eliminating γ effects, is

$$S_r = S_f - S_a = (X_{(2)}\beta+e)'(M-N)(X_{(2)}\beta+e) \ .$$

If we assume $E(b_ie_{ij}) = 0$, we have

$$\begin{aligned}
E(S_r) &= E(\beta'X'_{(2)}(M-N)X_{(2)}\beta) + E(e'(M-N)e) \\
&= tr(X'_{(2)}X(X'X)^*X'X_{(2)}-X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^*X'_{(1)}X_{(2)}) \\
&\qquad\qquad \cdot\sigma_b^2 + tr(M-N)\sigma_e^2 \\
&= tr(X'_{(2)}X_{(2)}-X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^*X'_{(1)}X_{(2)})\sigma_b^2+tr(M-N)\sigma_e^2 \ . \\
&= c\ \sigma_b^2 + (rk(M) - rk(N))\ \sigma_e^2
\end{aligned}$$

where

$$c = tr(X'_{(2)}X_{(2)} - X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^*X'_{(1)}X_{(2)}) \ .$$

Then $\hat{\sigma}_b^2 = \left\{ S_r - tr(M-N)/tr(I-M)\ Sa \right\} /c$

$= (S_r - d\, S_a)/c$, where $d = tr(M-N)/tr(I-M)$, and
$Var(\hat{\sigma}_b^2) = Var(S_r) + d^2\, Var(S_a) - 2d\, Cov(S_r S_a)/c^2$.

Let us consider obtaining $Var(S_r) = E(S_r^2) - (E(S_r))^2$

where S_r is the quadratic form $(X_{(2)}\beta+e)'(M-N)\,(X_{(2)}\beta+e)$. Let us denote the matrix of the form by $\{n_{ij}\}$, the i^{th} term of the vector $(X_{(2)}\beta+e)$ by $(X_{(2)}\beta+e)_i$, and limit ourselves to classification models, i.e., the elements of X are all 0 or 1 and there is only one non-zero term per row for $X_{(i)}$ matrices.

Accordingly, $X_{(2)}\beta$ can be strung out as a vector $(\dot{\beta}_1,\dot{\beta}_2,\ldots,\dot{\beta}_n)$. Now $S_r^2=(\sum_{ij}(n_{ij}\dot{\beta}_i\dot{\beta}_j+n_{ij}\dot{\beta}_i e_j+n_{ij}\dot{\beta}_j e_i+n_{ij}e_i e_j))^2$. Our interest does not center on expanding S_r^2 with all $\dot{\beta}_i$'s $(i=1,\ldots,n)$ distinct, but where $\dot{\beta}_i,\ldots,\dot{\beta}_{i+m}$ all equal β_i (say). So when we expand S_r^2, we shall regard $X_{(2)}\beta$ to be strung out as a vector containing repetitive terms, say,

$$(\beta_1,\ldots\beta_1,\beta_2,\ldots\beta_2,\beta_3,\ldots\beta_3,\ldots\beta_b) \qquad \text{(V.A.2)}$$

where the subvectors $(\beta_i,\ldots,\beta_i)$ are of dimensions determined by the structure of the experiment, the number of factors, repetitions, and so on.

The four top left-hand corner blocks of the full n x n matrix of a more complete representation of $\sum_{ij} n_{ij}\dot{\beta}_i\dot{\beta}_j$ and $\sum_{ij} n_{ij}\dot{\beta}_i e_j$ are written out below. It is assumed that the dimension of

$(\beta_1 \ldots \beta_1)'$ is t, and the dimension of $(\beta_2 \ldots \beta_2)'$ is t'.
We obtain

$$\left[\begin{array}{cccc|ccc}
n_{11}\beta_1^2 & n_{12}\beta_1^2 & \cdot\cdot & n_{1t}\beta_1^2 & n_{1,t+1}\beta_1\beta_2 & \cdot\cdot & n_{1,t+t'}\beta_1\beta_2 \\
\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\
\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\
n_{t1}\beta_1^2 & n_{t2}\beta_1^2 & \cdot\cdot & n_{tt}\beta_1^2 & n_{t,t+1}\beta_1\beta_2 & \cdot\cdot & n_{t,t+t'}\beta_1\beta_2 \\
\hline
n_{t+1,1}\beta_2\beta_1 & n_{t+1,2}\beta_2\beta_1 & \cdot\cdot & n_{t+1,t}\beta_2\beta_1 & n_{t+1,t+1}\beta_2^2 & \cdot\cdot & n_{t+1,t+t'}\beta_2^2 \\
\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\
\cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot \\
n_{t+t',1}\beta_2\beta_1 & n_{t+t',2}\beta_2\beta_1 & \cdot\cdot & n_{t+t',t}\beta_2\beta_1 & n_{t+t',t+1}\beta_2^2 & \cdot\cdot & n_{t+t',t+t'}\beta_2^2
\end{array}\right]$$

(V.A.3) ,

and

$$\left[\begin{array}{cc|cc}
n_{11}\beta_1 e_1 & n_{1t}\beta_1 e_t & n_{1,t+1}\beta_1 e_{t+1} & n_{1,t+t'}\beta_1 e_{t+t'} \\
n_{21}\beta_1 e_1 & n_{2t}\beta_1 e_t & n_{2,t+1}\beta_1 e_{t+1} & n_{2,t+t'}\beta_1 e_{t+t'} \\
\cdot & & & \\
n_{t1}\beta_1 e_1 & n_{tt}\beta_1 e_t & n_{t,t+1}\beta_1 e_{t+1} & n_{t,t+t'}\beta_1 e_{t+t'} \\
\hline
n_{t+1,1}\beta_2 e_1 & n_{t+1,t}\beta_2 e_t & n_{t+1,t+1}\beta_2 e_{t+1} & n_{t+1,t+t'}\beta_2 e_{t+t'} \\
n_{t+2,1}\beta_2 e_1 & n_{t+2,t}\beta_2 e_t & n_{t+2,t+1}\beta_2 e_{t+1} & n_{t+2,t+t'}\beta_2 e_{t+t'} \\
\cdot & \cdot & & \\
n_{t+t',1}\beta_2 e_1 & n_{t+t',t}\beta_2 e_t & n_{t+t',t+1}\beta_2 e_{t+1} & n_{t+t',t+t'}\beta_2 e_{t+t'}
\end{array}\right]$$

(V.A.4).

Ignoring the $\beta_i \beta_j$ terms in the representation (V.A.3) we represent the partition of $\{n_{ij}\}$ thereby indicated by

$$\begin{bmatrix} N_{11} N_{12} \cdot & \cdot & N_{1k} \\ N_{21} N_{22} \cdot & \cdot & \cdot \\ N_{k1} N_{k2} \cdot & \cdot & N_{kk} \end{bmatrix}_{n \times n} \qquad \text{(V.A.5)}$$

We have not written out a more complete representation for $\sum_{ij} n_{ij} \beta_j e_i$ because of the similarity between its representation and the one for $\sum_{ij} n_{ij} \beta_i e_j$. There is no need to write out a comparable representation for $\sum_{ij} n_{ij} e_i e_j$ because $E(e_i e_{i'}) = 0 (i \neq i')$.

In the representations (V.A.3) and (V.A.5) we note that the factor determining the subdivision is the definite dimension of each of the subvectors $(\beta_i, \ldots, \beta_i)'$ that comprise $X_{(2)}\beta$. When the two factors are not the same as in (V.A.4) the vertical subdivision is determined by e_i's, all of which are different, so the partition reflects this, while the horizontal partition is determined by β_j's. In later sections, when we consider the case of several random factors, we shall introduce the terminology of a β-partition (and a (β,ϵ)-partition) to distinguish possibly different subdivisions of $\{n_{ij}\}$ of the type (V.A.3) and (V.A.4) respectively, that are induced by different random factors (and sets of two random factors). Of course in the present case there is only one (β) subdivision that need concern us namely (V.A.5); we do not therefore really need to identify it as being the β-partition. Partitions involving e_i's will usually not be written down explicitly because of the special nature of the e vector.

Consider expanding the expression for (S_r^2). β_i^4 terms can arise by squaring any and all terms represented by N_{ii} in (V.A.5). There are doubtless numerous ways to represent the totality of such terms; we think the following both concise and informative. The terms in β_i^4 are given by $\sum_r N_r \beta_r^4$, where N_r is the sum of all terms of the matrix of direct products $N_{rr} \otimes N_{rr}$, where N_{rr}'s are diagonal blocks of partitioned matrices identified by (V.A.5), for example. Terms involving $\beta_r^2 \beta_s^2$ can be obtained in two ways. Included in the expansion of $E(S_r)^2$ are the two terms $\sum_{rs}^{\neq} M_{rs} \beta_r^2 \beta_s^2$ and $\sum_{rs}^{\neq} J_{rs} \beta_r^2 \beta_s^2$ where M_{rs} is two times the sum of all terms of the direct product $N_{rr} \otimes N_{ss}$ $(r \neq s)$, and assuming $\{n_{ij}\}$ is symmetric, J_{rs} is two times the sum of all terms of the direct product $N_{rs} \otimes N_{rs}$ $(r \neq s)$. The terms with non-zero expectation in e_i's are

$$\sum_i n_{ii}^2 e_i^4,\ 2\sum_{ij}^{\neq} n_{ii} n_{jj} e_i^2 e_j^2 \text{ and } 2\sum_{ij}^{\neq} n_{ij}^2 e_i^2 e_j^2 .$$

The terms in e_i's and β_j's are $\sum_{ir} L_{ir} \beta_r^2 e_i^2$ and $\sum_{ir} F_{ir} \beta_r^2 e_i^2$,

where the terms L_{ir} are two times the sum of all terms of the direct product $N_{rr} \otimes n_{ii}$, and where the terms F_{ir} are the sums of the matrix vector direct product

$$(\beta,e)\left\{\underset{\sim}{n}_{r_i}\right\} \otimes (\beta,e)\left\{\underset{\sim}{n}_{r_i}\right\} \qquad \text{(V.A.6)}$$

where by this notation $(\beta,e)\left\{n_{r_i}\right\}$ we mean to indicate that $\left\{\underset{\sim}{n}_{r_i}\right\}$ are in fact columns of the $\widetilde{N}_{rs}$ matrices and therefore have their dimension determined by these (N_{rs}) matrices and indirectly by the β-subdivision; their column partitioning (i.e. every column singly) is determined by e_i's.

To simplify expressions we use cumulants, described by Fisher (1930). The moment generating function of a probability law $\phi(x)$ is a function M defined for all real numbers t by

$$M = \int e^{tx}\,\phi(x)\,dx\ ,$$

where integration is over the domain of x. Then $K = \log M$ is the cumulant generating function of the probability law. K may be expanded in terms of moments, i.e., u_r's where $u_r = \int x^r\phi(x)\,dx$ or in terms of defined functions of u_r's denoted by K_i's or cumulants. Thus, for example, $K_1 = u_1$, $K_2 = u_2 - u_1^2$, $K_3 = u_3 - 3u_1u_2 + 2u_1^3$ and $K_4 = u_4 - 6u_2u_1^2 + 3u_2^2 + 8u_3u_1 - 6u_1^4$. When $u_1 = 0$, we have $K_1 = u_1 = 0$, $K_2 = u_2$, $K_3 = u_3$ and $K_4 = u_4 + 3u_2^2$.

In what follows the second subscript to a K_i will denote to which variable a particular cumulant refers. Thus K_{2i} denotes $E(e_i^2) - (Ee_i)^2 = \mathrm{Var}(e_i)$.

$$\text{Now } (ES_r)^2 = (\sum_i n_{ii}K_{2i} + \sum_{ij} n_{ij}E(\beta_i\beta_j))^2$$
$$= \sum_i n_{ii}^2(K_{2i})^2 + 2\sum_{ij}^{\neq} n_{ii}n_{jj}K_{2i}K_{2j}$$
$$+ \sum_{rs}^{\neq} M_{rs}K_{2\beta_r}K_{2\beta_s} + \sum_{ir} L_{ir}K_{2\beta_s}K_{2i} .$$

Therefore

$$\mathrm{Var}(S_r) = \sum_i n_{ii}^2(Ee_i^4 - (K_{2i})^2) + 2\sum_{ij}^{\neq} n_{ij}^2 K_{2i}K_{2j}$$
$$+ \sum_r N_r(E(\beta_r^4) - (K_{2\beta_r})^2) + \sum_{rs}^{\neq} J_{rs}K_{2\beta_r}K_{2\beta_s}$$
$$+ \sum_{ir} F_{ir}K_{2\beta_r}K_{2i} .$$

In cumulant notation we have

$$\mathrm{Var}(S_r) = K(S_r^2) = \sum_i n_{ii}^2 K_{4i} + 2\sum_{ij} n_{ij}^2 K_{2i}K_{2j}$$
$$+ \sum_r N_r K_{4\beta_r} + \sum_{rs} J_{rs}K_{2\beta_r}K_{2\beta_s} + \sum_{ir} F_{ir}K_{2\beta_r}K_{2i} .$$

Now

$$K(S_a^2) = K(\sum_{ij} m_{ij}e_ie_j)^2 \quad \text{where } \{mij\} = (I-M) .$$

By analogy with part of the above result we have

$$K(S_a^2) = \sum_i m_{ii}^2 K_{4i} + 2\sum_{ij} m_{ij}^2 K_{2i}K_{2j} .$$

We also have

$$K(S_aS_r) = \sum_i m_{ii}n_{ii}K_{4i} + 2\sum_{ij} m_{ij}n_{ij}K_{2i}K_{2j}$$
$$+ \sum_{ir} F^*_{ir}K_{2\beta_r}K_{2i} \quad \text{where } F_{ir}^*$$

is two times the sum of terms of the matrix vector direct product $(\beta, e)\{\underset{\sim}{m_{ri}}\} \otimes (\beta, e)\{\underset{\sim}{n_{ri}}\}$.

We note that if K_{2i} = a constant independent of i then $2\,\Sigma m_{ij} n_{ij} K_{2i}K_{2j} = 2K_{2i}K_{2j}\,tr(MN) = 0$. Under an assumption of normality $K_{4i} = 0$ and the final term of $K(S_r S_a)$ also vanishes because $E(\beta_r e_i\, \beta_r e_i) = E(\beta_r e_i)(E\beta_r e_i) = 0$.

In general, we have

$$Var(\hat{\sigma}_b^2) = \Big\{\sum_i n_{ii}K_{4i} + 2\sum_{ij} n_{ij}^2 K_{2i}K_{2j} + \sum_r N_r K_{4\beta_r} + \sum_{rs} J_{rs}K_{2\beta_r}K_{2\beta_s} + \sum_{ir} F_{ir}K_{2\beta_r}K_{2i} + d^2\Big(\sum_i m_{ii}^2 K_{4i} + 2\sum_{ij} m_{ij}^2 K_{2i}K_{2j}\Big) - 2d\Big(\sum_i m_{ii}n_{ii}K_{4i} + \sum_{ir} F_{ir}^* K_{2\beta_r}K_{2i}\Big)\Big\}/c^2 .$$

In the event that we can assume $K_{4i} = K_{4\beta_r} = 0$, some simplification is possible but the expression remains computationally speaking a rather formidable one:

$$\Big\{2\sum_{ij}(n_{ij}^2 + d^2 m_{ij})K_{2i}K_{2j} + \sum_{rs} J_{rs}K_{2\beta_r}K_{2\beta_s} + \sum_{ir}(F_{ir} - 2dF_{ir}^*)K_{2\beta_r}K_{2i}\Big\}/c^2 .$$

If e_{ij}'s are $N(0,\sigma^2)$ and b_i's are $N(0,\sigma_b^2)$,

$$V(\hat{\sigma}_b^2) = 1/c^2\Big\{2\sum_{ij}(n_{ij}^2 + d^2 m_{ij})\,\sigma^4 + \sum_{rs} J_{rs}\sigma_b^4\Big\} \qquad (V.A.6)$$

We shall later mention an alternative method of deriving (V.A.6).

c. <u>General forms for variances and covariances of quadratic forms in mixed models</u> We now extend the results of sub-sub-section b to the case of the variance of a quadratic form of the type

$$y'Ny=(X_{(1)}\gamma+X_{r+1}\beta+X_{r+2}\delta+\ldots+X_k\epsilon+e)'N(X_{(1)}\gamma+X_{r+1}\beta+\ldots+X_k\epsilon+e)$$

where by $X_{(1)}\gamma$ we represent the constant effects in the model and $X_{r+1}\beta$, $X_{r+2}\delta$, and so on are random elements. The first generalization over the results of the previous sub-sub-section that we make is therefore to accommodate several random factors. Secondly we shall regard N as a general matrix, and not necessarily require that N be a projection operator on some space of X_i matrices.

For reasons of convenience and manageability we restrict ourselves to classification experiment models. Then $(X_{r+1}\beta+\ldots+X_k\epsilon)$ can be rewritten

$$\begin{bmatrix}\beta_1\\ \beta_1\\ \beta_1\\ \cdot\\ \cdot\\ \cdot\\ \beta_t\end{bmatrix}+\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_1\\ \cdot\\ \cdot\\ \cdot\\ \delta_m\end{bmatrix}+\begin{matrix}\cdots\\ \cdots\\ \cdots\\ \\ \\ \\ \cdots\end{matrix}\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\\ \cdot\\ \cdot\\ \cdot\\ \epsilon_n\end{bmatrix}$$

where β_i is the effect of the i^{th} level of the first random factor represented by β, δ_i is the i^{th} level of the second random factor represented by δ and so on and of course the multiplicities of β_i and δ_i are completely unrelated to one another. In more general regression models it would be necessary to include functional parameter coefficients in the matrix $\{n_{ij}\}$ that would undergo partitioning.

Write $\{n_{ij}\}$ to represent the matrix of the quadratic form N and represent $X_{(1)}\gamma$ by a vector A_i.

In sub-sub section b, when considering only one random effect (apart from e) we had need of a "partitioning" of the matrix of the quadratic form (see for e.g. (V.A.5)) in accordance with the number of different values assumed by β and depending on the multiplicities of such values. We shall have reason to make use of various different "partitionings" according to β, δ,..., ε respectively in the present case, so it will be convenient to label a particular partitioning according to the vector to which it corresponds. Thus a β-partitioning of $\{n_{ij}\}$ refers to a blocking of rows and columns of $\{n_{ij}\}$ according to the multiplicities of different β_i values. Obviously this will sometimes require rearrangement of rows and columns of $\{n_{ij}\}$. There will be as many β-partitionings as there are random factors less one, since it is not necessary to write one out for e. We shall also have to contend with double partitionings of two types, namely (β, $\in$) and (β, e). We define the former later.

Now $K(S_r^2) = K(\sum_{ij} n_{ij}(\beta_i\beta_j + \beta_i\delta_j + \ldots + \beta_i\epsilon_j + \beta_i e_j + \beta_i A_j$

$$+\delta_i\beta_j + \delta_i\delta_j + \ldots + \delta_i\epsilon_j + \delta_i e_j + \delta_i A_j +$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$+ e_i\beta_j + \ldots + e_i\epsilon_j + e_i e_j + e_i A_j$$

$$+ A_i\beta_j + \ldots + A_i\epsilon_j + A_i e_j + A_i A_j))^2 .$$

By the use of an evaluation process fimiliar to the one followed in sub-sub-section b and in particular by making reference to the representations (V.A.3) and (V.A.5) it is apparent that instead of $\sum_r N_r K_{4\beta_r}$ we have $\sum_\beta \sum_r N_{r(\beta)} K_{4\beta_r}$ where summation β is over all β-partitions, and where $N_{r(\beta)}$ represents the sum of all terms of the matrix of direct products

$$(\beta)N_{rr} \otimes (\beta)N_{rr} .$$

Likewise instead of $\sum_{rs} J_{rs} K_{2\beta_r} K_{2\beta_s}$

we now have $\sum_\beta \sum_{rs} J_{rs(\beta)} K_{2\beta_r} K_{2\beta_s}$ where

$J_{rs(\beta)}$ is two times the sum of all terms of the direct product $(\beta)N_{rs} \otimes (\beta)N_{rs}$ $(r \neq s)$.

Terms arising from $\beta_i^2\ \delta_j^2$, which do not have a counterpart in sub-sub-section b, are given by $\overline{J}_{rs(\beta\delta)} K_{2\beta_r} K_{2\delta_s}$ where $\overline{J}_{rs(\beta\delta)}$ is two times the sum of all elements of the direct product of matrices $(\beta,\delta)M_{rs} \otimes (\beta,\delta)M_{rs}$; by $(\beta,\delta)M_{rs}$ we mean a horizontal partitioning according to β and a

vertical partitioning according to δ. See (V.A.6) for an example where we have already made use of this type of (double) partitioning. The term representing all product pairs of random elements of the above type is $\sum_{\beta\delta} \sum_{rs} \bar{J}_{rs(\beta\delta)} K_{2\beta_r} K_{2\delta_s}$.
The only term in β's and e's that need concern us, since the others will cancel, is the one that replaces $\sum_{ir} F_{ir} K_{2\beta_r} K_{2i}$ namely $\sum_{\beta} \sum_{ir} F_{ir(\beta)} K_{2\beta_r} K_{2i}$ where the terms $F_{ir(\beta)}$ are the sums of the direct product $(\beta,e)\{\underset{\sim}{n_{ri}}\} \otimes (\beta,e)\{\underset{\sim}{n_{ri}}\}$ where the vectors $\underset{\sim}{n_{ri}}$ are columns of $(\beta)N_{rs}$, i.e., the horizontal partitioning is according to the random parameter β, the vertical one according to e. Since the terms in K_{4i} and $K_{2i}K_{2j}$ will be the same as in the above sub-sub section we do not repeat them here.

Finally we must concern ourselves with the terms that involve the constant vector A_i and that enter the expression (S_r^2). These are $2\sum_{ij} m_{ii} m_{ij} A_j e_i^3$, $\sum_i (\sum_j m_{ij} A_j)^2 e_i^2$, $2\sum_{ri} D_{ri(\beta)} A_i \beta_r^3$ and $\sum_{ri} F_{ir(\beta)} A_i^2 \beta_r^2$; $D_{ri(\beta)}$ is the sum of terms of the direct product of matrix and vector $(\beta)N_{rr} \otimes (\beta,e)\{\underset{\sim}{n_{ri}}\}$, where the vectors $\{n_{ri}\}$ are columns of N_{rs}'s i.e. the column partitioning is according to the random parameter β, and where F_{ir} was previously defined.

The final result is:

$$K(S_r^2) = \sum_i m_{ii}^{\,2} K_{4i} + 2\sum_{ij} m_{ij}^{\,2} K_{2i} K_{2j} + \sum_\beta \sum_r N_{r(\beta)} K_{4\beta_r}$$

$$+ \sum_\beta \sum_{rs} J_{rs(\beta)} K_{2\beta_r} K_{2\beta_s} + \sum_\beta \sum_{ir} F_{ir(\beta)} K_{2\beta_r} K_{2i} + \sum_{\beta\delta} \sum_{rs} \bar{J}_{rs(\beta\delta)} K_{2\beta_r} K_{2\delta_s}$$

$$+ \sum_{ij} m_{ii} m_{ij} A_j K_{3i} + 2\sum_i \left(\sum_j m_{ij} A_j\right)^2 K_{2i}$$

$$+ \sum_\beta \sum_{ri} D_{ri(\beta)} A_i K_{3\beta_r} + \sum_\beta \sum_{ir} F_{ir} A_i^2 K_{2\beta_r} .$$

If we make the assumptions that all random effects have zero mean and their own constant respective variances, $K(S_r^2)$ is

$$2 \sum_{ij} m_{ij}^{\,2} K_{2i} K_{2j} + 2\sum_i \left(\sum_j m_{ij} A_j\right)^2 K_{2i} + \sum_\beta \sum_{ir} F_{ir(\beta)} K_{2i} K_{2\beta_r}$$

$$+ \sum_\beta \sum_{rs} J_{rs(\beta)} K_{2\beta_r} K_{2\beta_s} + \sum_\beta \sum_{ir} F_{ir} A_i^2 K_{2\beta_r} + \sum_{\beta\delta} \sum_{rs} \bar{J}_{rs(\beta\delta)} K_{2\beta_r} K_{2\delta_s}$$

(V.A.8).

We next consider obtaining a general form for the covariance of two quadratic forms of the type (say)

$$(X_{(1)}\gamma + X_{r+1}\beta + X_{r+2}\delta + \ldots + X_k\epsilon + e)' M (X_{(1)}\gamma + X_{r+1}\beta + X_{r+2}\delta + \ldots + X_k\epsilon + e)$$

and

$$(X_{(1)}\gamma + X_{r+1}\beta + \ldots\ldots + X_k\epsilon + e)' N (X_{(1)}\gamma + X_{r+1}\beta + \ldots\ldots + X_k\epsilon + e) .$$

The form of this covariance is the same as the variance formula above except that coefficient terms like $N_{r(\beta)}$, $J_{rs(\beta)}$ and so on are replaced by $N^*_{r(\beta)}$, $J^*_{rs(\beta)}$ etc. where starred terms arise from sums of direct products of matrices of the type

$(\beta)M_{rr} \otimes (\beta)N_{rr}$ and $(\beta,e)\{\underset{\sim}{m_{ri}}\} \otimes (\beta,e)\{\underset{\sim}{n_{ri}}\}$ for example, with similar restrictions to those that were placed on corresponding expressions involving only N_{rr} and $\{\underset{\sim}{n_{ri}}\}$ terms in the variance formula.

It can be shown that

$$\begin{aligned} K(S_r S_a) = {} & \sum_i m_{ii} n_{ii} K_{4i} + 2\sum_{ij} m_{ij} n_{ij} K_{2i} K_{2j} \\ & + \sum_{\beta}\sum_{r} M^*_{r(\beta)} K_{4\beta_r} + \sum_{\beta}\sum_{rs} J^*_{rs(\beta)} K_{2\beta_r} K_{2\beta_s} \\ & + \sum_{\beta}\sum_{ir} F^*_{ir(\beta)} K_{2\beta_r} K_{2i} + \sum_{\beta\delta}\sum_{rs} \bar{J}^*_{rs(\beta\delta)} K_{2\beta_r} K_{2\delta_s} \\ & + \sum_{ij} m_{ij} n_{ij} A_j K_{3i} + 2\sum_i \Big(\sum_j n_{ij} A_j\Big)\Big(\sum_j m_{ij} A_j\Big) K_{2i} \\ & + \sum_{\beta}\sum_{ir} D^*_{ir(\beta)} A_i K_{3\beta_r} + \sum_{\beta}\sum_{ri} B^*_{ri} K_{2\beta_r} \end{aligned} \qquad \text{(V.A.9).}$$

d. Special cases One of the advantages of the fitting constants method described in sub-section 1 in a mixed model comes about as a direct consequence of the way the method is constructed; we shall demonstrate firstly the considerable simplification of some of the formulae of sub-sub-section c that are possible with this method. In sub-sub-section e we give an alternative formula that is appropriate when normality is assumed.

In sub-section 1 we indicated that

$$\text{Rem}(\hat{\beta}_{r+1}) = y'(X(X'X)^*X' - \bar{X}(\bar{X}'\bar{X})^*\bar{X}')y$$

where

$$X = (X_0X_1\ldots X_rX_{r+1}\ldots X_k)$$

and

$$\bar{X} = (X_0X_1\ldots X_rX_{r+2}\ldots X_k) .$$

Write $X(X'X)^*X' = M$, $\bar{X}(\bar{X}'\bar{X})^*\bar{X}' = N$

Now $\text{Rem}(\beta_{r+1}) = \text{Rem}(\beta)$

$$=(X_{(1)}\gamma+X_{r+1}\beta+\ldots X_k\epsilon+e)'(M-N)(X_{(1)}\gamma+X_{r+1}\beta+\ldots+X_k\epsilon+e)$$

$$= (X_{r+1}\beta+e)'(M-N)(X_{r+1}\beta+e) .$$

We considered the variance of a form of this type in sub-sub-section b; in general then, variances of estimators of variance components in models that fit the requirements cause no problem. Although this is not usually specified, the extreme simplification found above will only arise in general in a classification model that does not involve interaction terms. We require some further specifications in order to deal with models of the latter type. One specification, for example, might be that when we fit ignoring an effect "A" say then we imply that interactions that involve "A" are to be ignored also. A residual S.S. that might then be obtained is (say) $\text{Rem}(\hat{\beta}_j,\ldots,\hat{\epsilon}_k)=(X_j\omega+\ldots X_k\epsilon+e)'(M-C)$

$$(X_j\omega+\ldots+X_k\epsilon+e) \qquad \text{(V.A.10)}$$

where $C = \dot{X}(\dot{X}'\dot{X})^*\dot{X}'$ and $\dot{X} = (X_0X_1.X_{j-1}X_{j+1}\ldots X_{k-1})$

For reasons of convenience and manageability we restrict ourselves to classification type models, wherein all the terms of the x_j matrices are either zero or one. Then $(X_j\omega+\ldots+X_k\epsilon+e)$ can be written

$$\begin{bmatrix}\omega_1\\ \omega_1\\ \omega_1\\ \omega_1\\ \omega_2\\ \cdot\\ \cdot\\ \omega_r\end{bmatrix} + \begin{bmatrix}\lambda_1\\ \lambda_1\\ \lambda_2\\ \lambda_2\\ \lambda_3\\ \cdot\\ \cdot\\ \lambda_m\end{bmatrix} + \begin{matrix}\cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\end{matrix} + \begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \epsilon_n\end{bmatrix}$$

where ω_i is the effect of the i^{th} level of the 1^{st} random factor represented by ω, λ_i is the i^{th} level of the second random factor represented by λ and so on, and of course the multiplicities of ω_i and λ_i are quite definite, but completely unrelated to one another.

The crucial point to note about this subdivision is that λ (say) is "nested" in ω in the sense that one or more levels of λ, will occur with a single level of ω. We attempted to emphasize this point in stringing out the individual vector components of $(X_j\omega+\ldots X_k\epsilon)$ above.

Write $\{m_{ij}\}$ to represent the matrix of the quadratic form $X(X'X)^*X'-\dot{X}(\dot{X}'\dot{X})^*\dot{X}' = (M-C)$. The similarity with the one matrix of the general form of the previous section is incidental. They are, needless to say, not necessarily the same. The required variance formula is then easily seen to be a simplified version of the general variance formula obtained in the previous section. In fact using similar notation, with the difference that partitionings are summed only over such factors as occur in the expression (V.A.10) we obtain $K(S_r^2) = \sum_i m_{ii}^2 K_{4i} + 2\sum_{ij} m_{ij}^2 K_{2i} K_{2j} + \sum\sum_{\omega r} N_{r(\omega)} K_{4\omega_r}$

$$+\sum_{\omega}\sum_{rs} J_{rs(\omega)} K_{2\omega_r} K_{2\omega_s} + \sum\sum_{\omega ir} F_{ir(\omega)} K_{2\omega_r} K_{2i} + \sum_{\omega\lambda}\sum_{rs} J_{rs(\omega\lambda)} K_{2\omega_r} K_{2\lambda_s} \quad .$$

(V.A.11)

The covariance of two different but overlapping quadratic forms of the type (V.A.10) with matrices "(M-C)" and "(M-D)" which for convenience we denote by M and N respectively is

$$K(S_r S_s) = \sum_i m_{ii} n_{ii} K_{4i} + 2\sum_{ij} m_{ij} n_{ij} K_{2i} K_{2j} + \sum_{\omega}\sum_{r} N^*_{r(\omega)} K_{4\omega_r}$$

$$+\sum_{(\omega)}\sum_{rs} J^*_{rs(\omega)} K_{2\omega_r} K_{2\omega_s} + \sum\sum_{\omega ir} F^*_{ir(\omega)} K_{2\omega_r} K_{2i} + \sum_{\omega\lambda}\sum_{rs} \overline{J}^*_{rs(\omega\lambda)} K_{2\omega_r} K_{2\lambda_s}$$

(V.A.12)

where the starred terms are similar to unstarred ones in form, however the direct products involve M_{rs} and N_{rs} matrices and such terms need only be considered when both quadratic forms involve similar random factors.

e. Variance and covariance of certain quadratic forms in normal variates The question arises whether we could obtain variances more easily by some other method, possibly by making more assumptions.

For convenient reference we restate first some results derived by other workers in the case where variates are assumed normally distributed which are probably easier to apply than those of sub-sub-sections a, b, and c for example. There are two major results for the following two cases:

i) y's are assumed to have zero (or constant) means and covariance V

ii) y's have non-zero mean and covariance V.

In case i) which is a special case of ii) we have, for positive definite matrices A and B,

$$\mathrm{Var}(y'Ay) = 2\ \mathrm{tr}(AV)^2 \quad \text{and}$$

$$\mathrm{Cov}(y'Ay, y'By) = 2\mathrm{tr}(AVBV).$$

An early reference to the latter result, which of course implies the former when A = B, is Matern (1949); it would come as no surprise however to hear that priority belonged to someone else. The forms for cumulants of y'Ay and joint cumulants of y'Ay, y'By .... are given by Lancaster (1954). Case ii) is implicit in a result stated, eg., by Plackett (1960), page 18. If y is multivariate normal (η, V) then the characteristic function of the quadratic form y'Ay is

$$\phi(t) = |I-2itAV|^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\eta' V^{-1}\eta + \tfrac{1}{2}\eta' V^{-1}(V^{-1}-2itA)^{-1}V^{-1}\eta\right\}$$

Theoretically we should be able to find all the moments of $y'Ay$ from this expression, however no one appears to have proceeded in this way.

In this section we relate the results of (V.A.11) and (V.A.12) to those mentioned in case (i).

Suppose that the quadratic form $\mathrm{Rem}(\beta_j,\ldots,\beta_k)$ is reducible to $z'(M-C)z$ where $E(z) = 0$ and $E(zz') =$

$$E(X_j\omega+\ldots+X_k\epsilon+e)(X_j\omega+\ldots+X_k\epsilon+e)' = Z.$$

Consequently $\mathrm{Var}(\mathrm{Rem}(\beta_j,\ldots,\beta_k)) = 2\ \mathrm{tr}((M-C)Z)^2$ and under normality we have a simple method for obtaining the counterpart of (V.A.11).

Let us consider a similar counterpart for (V.A.12). Let $\mathrm{Rem}(\beta_j,\ldots,\beta_k))$ and $\mathrm{Rem}(\beta_i,\ldots,\beta_1)$ be two overlapping S.S.'s. Suppose $\mathrm{Rem}(\beta_j,\ldots,\beta_k) = z'(M-C)z$ (as above) and $\mathrm{Rem}(\beta_i,\ldots,\beta_1) = w'(M-D)w$ (say) where $E(w) = 0$, and

$$E(ww') = E(X_i\rho+\ldots+X_1\tau+e)(X_i\rho+\ldots+X_1\tau+e)' .$$

We have given the general formula for $\mathrm{Cov}(\mathrm{Rem}(\beta_j,\ldots,\beta_k),$ $\mathrm{Rem}(\beta_i,\ldots,\beta_1))$, i.e., (V.A.12). We now show that if we assume normality, the following more compact formulation may be used to determine the required covariance. The covariance between two quadratic forms, one in z, distributed as $N(0,Z)$, the other in w, distributed as $N(0,W)$, where z and w are vectors of the same order, is $\mathrm{Cov}(z'Qz, w'Pw) = 2\mathrm{tr}(QZPW)$

(V.A.13).

We prove this by showing that (V.A.13) gives the same result term by term as does the more general form (V.A.12), when K's of order greater than two in the latter form are taken to be zero.

We have $Z = E(X_j\omega+\ldots+X_k\epsilon+e)(X_j\omega+\ldots+X_k\epsilon+e)'$. The elements of Z that are contributed by different components (i.e. $X_j\omega$'s) are of a special "nested" nature. Thus all terms contributed by $X_k\epsilon$ occur in positions that already have a contribution. In the representation of the four top left-hand corner blocks of Z, determined by the ω-partition, we try to emphasize this point. Thus $E(X_j\omega)(X_j\omega)'$ will give rise to

$$\left[\begin{array}{cccc|cccc}
K_{2\omega_1} & K_{2\omega_1} & \cdots & K_{2\omega_1} & & & & \\
K_{2\omega_1} & K_{2\omega_1} & \cdots & K_{2\omega_1} & & 0 & & \\
\cdot & \cdot & \cdot & \cdot & & & & \\
K_{2\omega_1} & K_{2\omega_1} & \cdots & K_{2\omega_1} & & & & \\
\hline
 & & & & K_{2\omega_2} & K_{2\omega_2} & \cdots & K_{2\omega_2} \\
 & & & & K_{2\omega_2} & K_{2\omega_2} & \cdots & K_{2\omega_2} \\
 & 0 & & & \cdot & \cdot & \cdot & \cdot \\
 & & & & K_{2\omega_2} & K_{2\omega_2} & \cdots & K_{2\omega_2}
\end{array}\right] .$$

$E(X_{j+1}\lambda)(X_{j+1}\lambda)'$ contributions will not fill out the top left hand block, but will also not give terms in blocks that do not already have non-zero elements. We indicate this as follows:

$$
\left[\begin{array}{ccccc|cccc}
K_2\lambda_1 & K_2\lambda_1 & \cdot & 0 & 0 & & & & \\
K_2\lambda_1 & K_2\lambda_1 & \cdot & 0 & 0 & & & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & & & & \\
0 & 0 & \cdot & K_2\lambda_2 & K_2\lambda_2 & & & & \\
0 & 0 & \cdot & K_2\lambda_2 & K_2\lambda_2 & & & & \\
\hline
 & & & & & K_2\lambda_j & K_2\lambda_j & & \\
 & & & & & K_2\lambda_j & K_2\lambda_j & & \\
 & & & & & & \cdot \quad \cdot & \cdot & \cdot \\
 & & & & & & & K_2\lambda_k & K_2\lambda_k \\
 & & & & & & & K_2\lambda_k & K_2\lambda_k
\end{array}\right]
$$

We note that further contributions will satisfy similar "nesting" conditions, and in particular, the last element E(ee') gives rise to

$$
\begin{bmatrix}
K_{21} & & & & & & & \\
 & K_{22} & & & & & & \\
 & & \cdot & & & & & \\
 & & & \cdot & & & & \\
 & & & & \cdot & & & \\
 & & & & & \cdot & & \\
 & & & & & & \cdot & \\
 & & & & & & & K_{2n}
\end{bmatrix}
$$

We can now formally obtain QZ. Only the first four top left-hand corner blocks are given. We give the contributions separately and in the order that these are given by individual component (matrices) of Z.

We obtain firstly

$$\left[\begin{array}{ccc|cc}
(\omega)N_{11\underline{1}}K_2\omega_1 & (\omega)N_{11\underline{1}}K_2\omega_1 \cdot\cdot & (\omega)N_{11\underline{1}}K_2\omega_1 & (\omega)N_{12\underline{1}}K_2\omega_2 \cdot\cdot & (\omega)N_{12\underline{1}}K_2\omega_2 \\
(\omega)N_{11\underline{2}}K_2\omega_1 & (\omega)N_{11\underline{2}}K_2\omega_1 \cdot\cdot & (\omega)N_{11\underline{2}}K_2\omega_1 & (\omega)N_{12\underline{2}}K_2\omega_2 \cdot\cdot & (\omega)N_{12\underline{2}}K_2\omega_2 \\
\cdot & \cdot\cdot & & \cdot\cdot & \\
(\omega)N_{11\underline{t}}K_2\omega_1 & (\omega)N_{11\underline{t}}K_2\omega_1 \cdot\cdot & (\omega)N_{11\underline{t}}K_2\omega_1 & (\omega)N_{12\underline{t}}K_2\omega_2 \cdot\cdot & (\omega)N_{12\underline{t}}K_2\omega_2 \\
\hline
(\omega)N_{21\underline{1}}K_2\omega_1 & (\omega)N_{21\underline{1}}K_2\omega_1 \cdot\cdot & (\omega)N_{21\underline{1}}K_2\omega_1 & (\omega)N_{22\underline{1}}K_2\omega_2 \cdot\cdot & (\omega)N_{22\underline{1}}K_2\omega_2 \\
 & \cdot\cdot & & & \\
 & \cdot\cdot & & & \\
(\omega)N_{21\underline{t}}K_2\omega_1 & (\omega)N_{21\underline{t}}K_2\omega_1 \cdot\cdot & (\omega)N_{21\underline{t}}K_2\omega_1 & (\omega)N_{22\underline{t}}K_2\omega_2 \cdot\cdot & (\omega)N_{22\underline{t}}K_2\omega_2
\end{array}\right]$$

where by $(\omega)N_{ij\underline{k}}$ we mean the sum of the k^{th} row of the symmetric matrix N_{ij} when the partitioning is according to ω.

Next we obtain the terms that are contributed by the second matrix of Z above. In view of the nature of that matrix, we point out that the λ-partioning of N gives a "smaller" grid of N_{ij} terms than does the ω-partitioning, i.e. terms with $(\lambda)N_{11\underline{1}}$ coefficient do not extend over the whole first block. The following representation of the contribution to the top four blocks of QZ is meant to reflect this. We obtain

$$
\left[\begin{array}{cc|c|c|c}
(\lambda)N_{11\underline{1}}K_2\lambda_1 \cdot & (\lambda)N_{11\underline{1}}K_2\lambda_1 & (\lambda)N_{12\underline{1}}K_2\lambda_2 & (\lambda)N_{12\underline{1}}K_2\lambda_2 \cdot & (\lambda)N_{13\underline{1}}K_2\lambda_3 \\
\lambda N_{11\underline{s}}K_2\lambda_1 \cdot & (\lambda)N_{11\underline{s}}K_2\lambda_1 & (\lambda)N_{12\underline{s}}K_2\lambda_2 & (\lambda)N_{12\underline{s}}K_2\lambda_2 \cdot & (\lambda)N_{13\underline{s}}K_2\lambda_3 \\
\hline
\lambda N_{21\underline{1}}K_2\lambda_1 \cdot & (\lambda)N_{21\underline{1}}K_2\lambda_1 & (\lambda)N_{22\underline{1}}K_2\lambda_2 & (\lambda)N_{22\underline{1}}K_2\lambda_2 \cdot & (\lambda)N_{23\underline{1}}K_2\lambda_3 \\
\hline
\lambda N_{21\underline{2}}K_2\lambda_1 \cdot & (\lambda)N_{21\underline{2}}K_2\lambda_1 & (\lambda)N_{22\underline{2}}K_2\lambda_2 & (\lambda)N_{22\underline{2}}K_2\lambda_2 \cdot & (\lambda)N_{23\underline{2}}K_2\lambda_3 \\
\\
\hline
(\lambda)N_{31\underline{u}}K_2\lambda_1 \cdot & (\lambda)N_{31\underline{u}}K_2\lambda_1 & (\lambda)N_{32\underline{u}}K_2\lambda_2 & (\lambda)N_{32\underline{u}}K_2\lambda_2 \cdot & (\lambda)N_{33\underline{u}}K_2\lambda_3
\end{array}\right]
$$

Similarly we could write out the matrix contributions for all the remaining random factors. Finally we obtain the complete contribution from E(ee') as follows.

$$
\begin{bmatrix}
n_{11}\ K_{21} & n_{12}\ K_{22} & & n_{1n}\ K_{2n} \\
n_{21}\ K_{21} & n_{22}\ K_{22} & & n_{2n}\ K_{2n} \\
 & & & \\
n_{n1}\ K_{21} & n_{n2}\ K_{22} & & n_{nn}\ K_{2n}
\end{bmatrix}
$$

The form of PW we shall not write out in detail since it so closely resembles in basic structure the one exhibited above for QZ. Of course the factors entering PW are not necessarily either the same or different from the ones entering QZ. Partitions may therefore be different or the same. The important point to note is that partitions of W are "nested" in earlier occurring ones in the same way as this was the case in Z.

Let us compare coefficients for like terms in (V.A.13) and in the general formula (V.A.12) above. Consider first the coefficient of $K_{2i}K_{2j}$ in tr(QVPW) .

We find the terms involving K_{2i}'s and K_{2j}'s to be

$$n_{11}m_{11}K_{21}^2+n_{12}m_{21}K_{22}+n_{13}m_{31}K_{21}K_{23}+\cdots$$

$$n_{21}m_{12}K_{21}K_{22}+n_{22}m_{22}K_{22}^2+n_{23}m_{32}K_{23}K_{22}+\cdots$$

$$n_{31}m_{13}K_{21}K_{23}+n_{32}n_{23}K_{22}K_{23}+n_{33}m_{33}K_{23}\cdots$$

$= \sum_{ij} n_{ij}m_{ij}K_{2i}K_{2j}$, so that when multiplied by 2, this term agrees exactly with the corresponding term in (V.A.12).

Consider next the $K_{21}K_{2\omega_1}$ term. We find this to be

$$^{(\omega)}N_{11\underline{1}}m_{11}K_{2\omega_1}K_{21}+{}^{(\omega)}N_{11\underline{1}}m_{21}K_{2\omega_1}K_{21}+\cdots{}^{(\omega)}N_{11\underline{1}}m_{t1}K_{2\omega_2}K_{21}$$

$$= {}^{(\omega)}N_{11\underline{1}} \otimes {}^{(\omega)}M_{11\underline{1}} = F^{*}_{11(\omega)}$$

in agreement with the coefficient found in (V.A.12) above. By summing over i and r and over ω we take care of $K_{2i}K_{2\eta_r}$, $K_{2i}K_{2\varepsilon}$, and so on.

Finally we come to terms of the type $K_{2\omega_r}K_{2\lambda_s}$, etc. The coefficient of $K_{2\omega_1}K_{2\lambda_1}$ for example is

$$^{(\omega)}N_{11\underline{1}}\cdot{}^{(\lambda)}M_{11\underline{1}}+{}^{(\omega)}N_{11\underline{1}}\cdot{}^{(\lambda)}M_{11\underline{2}}+\cdots+{}^{(\omega)}N_{11\underline{1}}\cdot{}^{(\lambda)}M_{11\underline{t}}$$

$$^{(\omega)}N_{11\underline{2}}\cdot{}^{(\lambda)}M_{11\underline{1}}+{}^{(\omega)}N_{11\underline{2}}\cdot{}^{(\lambda)}M_{11\underline{2}}+\cdots+{}^{(\omega)}N_{11\underline{2}}{}^{(\lambda)}M_{11\underline{t}}$$

$$+\cdots\cdots+\cdots\cdots+{}^{(\omega)}N_{11\underline{t}}{}^{(\lambda)}M_{11\underline{t}}$$

$$= (\omega,\lambda)N_{11} \cong (\omega,\lambda)M_{11} = J^*_{11(\omega\lambda)},$$

where the partitioning (ω,λ) is defined in sub-sub-section b. Of course if ω and λ are the same, there is no need for a double partitioning and we have the coefficient $J^*_{11}(\omega)$ for $K^2_{2\omega_1}$, which was also defined in sub-sub-section b. By summing over r and s, ω and λ we take care of terms of the type $K_{2\omega_r}K_{2\lambda_s}$ and $K_{2\omega_r}K_{2\rho_s}$ etc. Agreement between the form given by (V.A.12) (modified) and the form (V.A.13) is therefore complete.

B. Applications to Covariance Type Models

1. Random models with concomitants

A sub-class of situations falling within the general class of a mixed unbalanced model has received some attention by Crump (1947, 1951). This sub-class refers to covariance models with classification terms random (rather than fixed). Crump (loc.cit.) gave two methods by which we might estimate variance components in models of this type. Crump limited his discussion to a one-way classification with a single concomitant. The two methods that he suggested do however generalize quite easily to more complex cases. Crump made no explicit mention of the general least squares method, but one of the methods he suggested was the conventional A.o.C. procedure for the corresponding case when all effects are fixed, to give right-hand sides of equations, the left-hand sides of which are expectations in accord with the actual model. As is well known, this is equivalent to the L.S. procedure.

The general form of the model presently envisaged (in conventional notation) is

$$y_{ij} = \mu + \gamma_1 x_{ij}^{(1)} + \ldots + \gamma_m x_{ij}^{(m)} + a_i + b_j + \ldots + e_{ij}$$

where γ_i's are fixed unknown parameters, $x_{ij}^{(m)}$'s are fixed known numbers, and a_i's, b_j's,...,e_{ij}'s are random parameters assumed independent of each other. To bring this more in line with the notation of this thesis, we represent the model above by $y = X_{(1)}\gamma + X_{(2)}\beta + e$ where γ refers to the fixed γ_i parameters, and β on the other hand refers to all the terms (i.e., both fixed and random) of the model without concomitants.

The general form of an A.o.V. sub-division (with only one concomitant x shown) is

Source	S.S. y	S.P.$_{xy}$	S.S.$_x$
β_1	A_{yy}	A_{xy}	A_{xx}
β_2	B_{yy}	B_{xy}	B_{xx}
.	C_{yy}	C_{xy}	C_{xx}
Error	E_{yy}	E_{xy}	E_{xx}
(β_1 + error)	$(A+E)_{yy}$	$(A+E)_{xy}$	$(A+E)_{xx}$

and by A_{yy}, for example, we mean Rem(β_1)(=Rem(a_i)) i.e. the difference between the S.S. due to fitting all effects and the S.S. due to fitting all but a_i's in a model without concomitants. Let us represent the model (given above) when all β_1(=a_i) effects are put equal to zero by

$$y = X_{(1)}\gamma + \bar{X}_{(2)}\beta + e$$

then

$$A_{yy} = y'(X_{(2)}(X'_{(2)}X_{(2)})^*X'_{(2)} - \bar{X}_{(2)}(\bar{X}'_{(2)}\bar{X}_{(2)})^*\bar{X}'_{(2)})\ y$$

$$= y'(M - N)\ y$$

$$= y'(M - N)'(M - N)\ y$$

where $M = X_{(2)}(X_{(2)}'\ X_{(2)})^*\ X'_{(2)}$ and

$$N = \bar{X}_{(2)}(\bar{X}_{(2)}'\bar{X}_{(2)})^*\bar{X}'_{(2)}\ .$$

Now $A_{xy} = X'_{(1)}(M-N)y$, by simple analogy with A_{yy}, and in the same way $A_{xx} = X'_{(1)}(M-N)\, X_{(1)}$.

In the same way we find

$$E_{yy} = y'(I - X_{(2)}(X'_{(2)}\, X_{(2)})^{*}X'_{(2)})y = y'(I-M)y$$

so that $E_{xy} = X'_{(1)}(I - M)\, y$ and

$$E_{xx} = X'_{(1)}(I - M)\, X_{(1)} \; .$$

We shall make use of the accepted two-stage procedure for finding adjusted mean squares in the Model I analysis of covariance without elaboration.

Thus, the minimum S.S. in the model with concomitants is $SS_E = E_{yy}-E'_{xy}E^{*}_{xx}E_{xy} = y'(I-M)y-\hat{\gamma}'X'_{(1)}(y-X_{(2)}\tilde{\beta})$ where

$\beta = (X'_{(2)}X_{(2)})^{*}X'_{(2)}y$ and $\hat{\gamma}$ is a solution to the equations

$$X'_{(1)}\ (I-M)\ X_{(1)}\gamma = X'_{(1)}(I-M)\ y,$$

i.e. $E_{xx}\gamma = E_{xy}$.

Therefore $\hat{\gamma} = (X'_{(1)}(I-M)\, X_{(1)})^{*}\, X'_{(1)}(I-M)_{y}$.

By substitution we obtain

$$\begin{aligned} E_{yy}-E'_{ry}\, E^{*}_{xx}E_{xy} &= y'((I-M)-(I-M)X_{(1)}(X'_{(1)}(I-M)X_{(1)})^{*}X'_{(1)}(I-M))y \\ &= (X_{(1)}\gamma+X_{(2)}\beta+ e)'R(X_{(1)}\gamma+X_{(2)}\beta+e) \\ &= e'Re \end{aligned}$$

where $R = \{r_{ij}\} = ((I-M)-(I-M)X_{(1)}(X'_{(1)}(I-M)X_{(1)})^{*}X'_{(1)}(I-M))$.

We therefore have a general form for

$$E(SS_E) = tr(RV) = \sum_i r_{ii} K_{2i}$$

and by the results of sub-section 2

$$Var(SS_E) = \sum_i r_{ii}K_{4i} + 2\sum_{ij} r_{ij}^2 K_{2i}K_{2j} \quad (V.A.14) .$$

Alternatively if we can assume normality of error terms we may write

$$(Var(SS_E) = 2\ tr(RV)^2 .$$

We have exhibited a general variance formula for the adjusted error line of a model that has concomitants added.

Further manipulations are closely analogous to those above. We only have to bear in mind the matrices of the quadratic form. We find estimators for σ_A^2 first by Procedure one (of Crump (1947)), then by Procedure two.

Thus $SS_A = A_{yy}-A'_{xy}(A_{xx})^{*} A_{xy}$

$$= y'(M-N)-(M-N)X_{(1)}(X'_{(1)}(M-N)X_{(1)})^{*}X'_{(1)}(M-N)),$$

Then $E(SS_A) = tr(QV)$ where V, the covariance matrix of y, can be reduced to resemble Z (or W) of sub-section 2, section A.

Consequently

$$\begin{aligned}\mathrm{Var}(SS_A) = &\sum_i q_{ii}^2 K_{4i} + \sum_{ij} q_{ij}^2 K_{2i}K_{2j} + \sum_{\omega}\sum_{r} N^+_{r(\omega)} K_{4\omega_r} \\ &+ \sum_{\omega}\sum_{rs} J^+_{rs(\omega)} K_{2\omega_r} K_{2\omega_s} + \sum_{\omega}\sum_{ir} F^+_{ir(\omega)} K_{2\omega_r} K_{2i} \\ &+ \sum_{\omega\lambda}\sum_{rs} J^+_{rs(\omega\lambda)} K_{2\omega_r} K_{2\lambda_s} ,\end{aligned}$$

where daggers denote terms analogous to those without daggers in (V.A.11), and are derived using a matrix $Q=\{q_{ij}\}$ instead of $M = \{m_{ij}\}$.

If an assumption of normality of random effects is in order we have $\mathrm{Var}(SS_A) = 2tr(QV)^2$ (V.A.15) .

Finally we obtain similar expressions for S.S.'s, obtained by Procedure 2.

$$\begin{aligned}SS_{A+E} &= (A+E)_{yy} - (A+E)'_{xy}\,(A+E)^*_{xx}\,(A+E)_{xy} \\ &= y'(I-N)y - \tilde{\gamma}'\,X'_{(1)}\,(y - X_{(2)}\tilde{\tilde{\beta}})\end{aligned}$$

where $\tilde{\gamma}$ is a solution to the equations

$$X'_{(1)}(I-N)\,X_{(1)}\gamma = X'_{(1)}(I-N)y \text{ , and}$$

$\tilde{\tilde{\beta}} = (\bar{X}'_{(2)}\bar{X}_{(2)})^*\bar{X}'_{(2)}\,y$. Therefore

$$\begin{aligned}SS_{A+E} &= y'((I-N)-(I-N)X_{(1)}(X'_{(1)}(I-N)X_{(1)})^*X'_{(1)}(I-N))y \\ &= y'Py ,\end{aligned}$$

where

$$P = \{p_{ij}\} = ((I-N)-(I-N)X_{(1)}(X'_{(1)}(I-N)X_{(1)})^{*}X'_{(1)}(I-N))$$

Now $ASS_A = y'(P-R)y = y'Sy$. Then $E(ASS_A) = tr(SV)$

and

$$Var(ASS_A) = \sum_i s_{ii}^2 K_{4i} + \sum_{ij} s_{ij}^2 K_{2i}K_{2j} + \sum_{\beta}\sum_{r} N_r^{**} K_{4\beta_r} +$$

$$+ \sum_{\beta}\sum_{rs} J_{rs(\beta)}^{**} K_{2\beta_r}K_{2\beta_s} + \sum_{\beta}\sum_{ir} F_{ir}^{**} K_{2\beta_r}K_{2i}$$

$$+ \sum_{\beta\delta}\sum_{rs} \bar{J}_{rs(\beta\delta)}^{**} K_{2\beta_r}K_{2\delta_s}$$

where ** denotes terms analogous to those without **'s in (V.A.11), and which are derived using a matrix $S = \{s_{ij}\}$ instead of $M = \{m_{ij}\}$.

If an assumption of normality of random effects is in order, we have $Var(ASS_A) = 2\, tr(SV)^2$ (V.A.16) . The forms (V.A.14), (V.A.15) and (V.A.16) are generalizations of corresponding formula exhibited by Crump (1947) for a one-way classification with one concomitant, to any number of concomitants and higher-way classifications that do not necessarily have equal numbers in the cells.

We could also write down formulae for covariance (SS_A, SS_B) for example, along the lines described in Section A.

An overall measure of effectiveness of the two methods may then be obtained in the way described in Chapter IV.

Thus, we may obtain

$$\begin{bmatrix} \hat{\sigma}^2_{r+1} \\ \hat{\sigma}^2_{r+2} \\ \cdot \\ \cdot \\ \hat{\sigma}^2_{k+1} \end{bmatrix} = U^{-1} \begin{bmatrix} SS_A \\ SS_B \\ \\ \\ SS_E \end{bmatrix}$$

and the covariance matrix of the vector $(\hat{\sigma}^2_{r+1},\ldots,\hat{\sigma}^2_{k+1})'$ could be written down. We may do the same thing for estimators given by procedure two. We could then obtain an overall measure of the effectiveness of a procedure by calculating determinant or trace of the covariance matrix of the estimators, and comparing different procedures.

2. <u>Estimators of variance components when missing observations occur in designed situations</u>

This chapter would not be complete without some statement on how the results of section A relate to the missing value problem. We describe the fixed effects case first, where a frequently used method of adjusting for missing values is the covariance method.

Covariance procedures in experimental design are, it would seem, most readily understood as a device for including additional parameters in a model which can be easily fitted, and, as a consequence, as a systematic missing value procedure. The A.o.C. was introduced by Fisher in 1932. Scheffè's (1959) description of the procedure is that the A.o.C. is "a device for simulating control of factors not possible or feasible to control in the experiment; thus the estimates of the yields of varieties of grain in a comparative agricultural trial might be "adjusted" to allow for differing numbers of shoots on the plots or the plot's yields on a previous year's uniformity trial (where all plots are given the same treatment), and the resulting estimates would within sampling errors be the same as those that would be obtained if all plots had equal numbers of shoots or equal yields on the uniformity trial -- <u>granting the validity of the regression model assumed or implied</u>." (Our emphasis.) This view of the A.o.C. as a means for increasing precision of desired comparisons is of course implied by our statement above. Usually controversy in the A.o.C. is in the area of interpretation, and largely because it is seldom true that the exacting mathematical restrictions implicit in the technique are applicable to real data. Insofar as the assumptions are approximately true the technique is useful.

When making all the statements above, one invariably has only Model I situations in mind. In sub-section 1 of this section we discussed some methods of estimation in a general A.o.C. situation in which the aim is to extend the scope of the analysis of covariance to "mixed" models. We notice, of course, that interpretation is likely to cause even more trouble than before, but apart from this, the extension appears to be in order.

We now describe in brief the use of the covariance method in Model I missing value situations. The classical missing plot problem envisages cases where several of the planned observations are missing. It is assumed that such "missing observations" were not caused by treatments, and their occurrence is in fact beyond the control of the experimenter. From one point of view we might regard the remaining observations as forming an unbalanced design, and analyze accordingly by least squares ab initio.

Alternatively, the planned observations would have been represented by the model

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \gamma + e \qquad \text{(V.B.1)}$$

so that it is convenient to represent the available observations by

$$\begin{aligned} y_2 &= W_2 \gamma + e \qquad \text{(V.B.2)} \\ &= W_3 \gamma_1 + W_4 \gamma_2 + e \text{ (say)} . \end{aligned}$$

(We have introduced W's with subscripts instead of more conventional X's to avoid confusion with the use of X_i's throughout this thesis.) The covariance method provides us with a useful means of calculating min. S.S. H_j and estimates for (V.B.2). The method was first suggested by

Bartlett (1937) and it requires augmenting the observation vector with zeros for responses that are not available and artificially introducing concomitants, in a special well-known way. If we perform these recommendations we obtain as a new representation of (V.B.2)

$$\begin{bmatrix} 0 \\ y_2 \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \gamma + Z\delta + e \qquad \text{(V.B.3)}$$

where $Z = \begin{bmatrix} -I \\ 0_1 \end{bmatrix}$,

0 is a m x 1 vector of zeros,

I is a m x m matrix, and

0_1 is a n-m x m matrix of zero elements.

The reason for going to the model (V.B.3) when (V.B.2) is the situation, is based firstly on the desire to be able to make use of the systematic methods of dealing with (V.B.3) to which we have made mention in sub-section 1 and secondly because it is possible to choose the augmentations in such a way that min. S.S. H_j and estimates as obtained from models (V.B.2) and (V.B.3) will agree.

Certain pertinent points may now be mentioned.

a) The main reason for the development of the covariance method of dealing with missing values originally was to simplify the analysis. At that time the inversion of large matrices was an arduous task. Missing value techniques, among them the covariance technique, helped one circumvent this difficulty to a considerable extent.

b) Opinions will vary on whether such short cut methods should be retained when least squares ab initio presents no difficulty to a modern electronic machine. This author is of the opinion that since the A.o.C. routines will probably be retained in computer libraries along with the A.o.V., dealing with missing value situations by the A.o.C. method will be fast and efficient.

In Model II situations, comparable simplifying procedures do not appear to have been discussed. The fairly extensive literature on similar procedures in Model 1 situations makes this surprising.

In view of the use of the least squares approach (in this thesis) which requires the postulation of fixed effects, to find estimators in the unbalanced case, it can be shown that the A.o.C. method may be used as a simplifying procedure for at least part of the way. Thus, in setting up the equations that are to be solved for point estimators of the variance components, $\mathrm{Rem}(\beta_i)$ (say), values are required. These may be obtained in the conventional way for a fixed effects model by use of the A.o.C. procedure. However, in obtaining the $E(\mathrm{Rem}(\beta_i))$ values it is suggested that we apply the method spelled out in Chapter IV as this applies to the original model. We may use the variance formulae of subsection one to evaluate the precision of the least squares method of estimation.

We note that when a large proportion of the data is missing, it will probably be more economical to use a straight least squares approach on the available data. Finally we point out that the solution obtained by least squares is only one method among many that are possible; more work needs to be done on competitors of unknown performance.

VI. SUMMARY

In this thesis we consider the problem of minimum variance (M.V.) unbiased estimation of regression parameters and variance components in the mixed model

$$y = \sum_{i=0}^{r} X_i \gamma_i + \sum_{i=r+1}^{k+1} X_i \beta_i \qquad \text{(VI.1)} ,$$

where γ_i's are fixed effects, β_i's are random effects with distributional properties to be further specified, and X_i's are known fixed matrices whose elements are not necessarily restricted to be 0's or 1's. We assume throughout that $X_{k+1} = I$, $E(\beta_i \beta_j') = 0$ $(i \neq j)$, and $E(\beta_{k+1} \beta_{k+1}') = I\, \sigma_{k+1}^2$.

Two results for the model

$$y = j_n \mu + \sum_{i=1}^{k+1} X_i \beta_i = \sum_{i=0}^{k+1} X_i \beta_i$$

$$\text{where } E(\beta_i \beta_i') = I\, \sigma_i^2 \qquad \text{(VI.2)}$$

which are due to Graybill and Hultquist (1961), and which we have refined in this thesis are:

1. If a) all σ_i^2 are estimable b) $X_i X_i' X_j X_j' = X_j X_j' X_i X_i'$ $(i,j=0,\ldots,k+1)$ and c) the random β_i vectors are normally distributed then there is a complete sufficient statistic for the parameters $(\mu, \sigma_i^2, \ldots, \sigma_{k+1}^2)$ if, and only if, $W = \sum_{i=1}^{k+1} X_i X_i' \sigma_i^2 + X_0 X_0' \mu^2$ has k+2 distinct latent roots. The

set of sufficient statistics consists of $\overline{y}$ and $y'P_i'P_iy/c_i$ $(i=1,\ldots,k+1)$ where P_i's are collections of vectors of P, an orthogonal matrix such that $PWP' = \Delta$ (diagonal), and where all vectors of P_i (say) correspond to the same latent root of W.

We define the class of situations of type (VI. 2) for which commutativity of $X_iX_i'X_jX_j'$ $(i,j=0,\ldots,k+1)$ holds and W has k+2 distinct roots to be the class P.

2. Suppose we consider only those cases within the class P for which it is true that the diagonal submatrices of specified order of P_iP_i' are equal. If all fourth moments exist for all random variables, and all third moments are equal and all fourth moments are equal for the elements of a given β_i vector, then the same estimators for variance components that are M.V. unbiased under normality are best quadratic unbiased under the assumptions mentioned.

We have obtained analogous results to 1, and under slightly more extended restrictions to 2 above, for the completely random model under the assumption that $E(\beta_i\beta_i') = (a_i \setminus b_i)$ $(i=1,\ldots,k)$ where $a_i \setminus b_i$ is a matrix with a_i on the diagonal and b_i off it. The same estimators as before are complete sufficient for $(\mu, a_1-b_1,\ldots a_k-b_k, \sigma_{k+1}^2)$.

For the mixed model (VI. 1), under the assumptions a) of normality of β_i vectors b) $E(\beta_i\beta_i') = I\,\sigma_i^2$ $(i=r+1,\ldots,k+1)$ c) $X_iX_i'X_jX_j' = X_jX_j'X_iX_i'$ $(i,j=0,\ldots,k+1)$

d) $\overline{W}$ has k+2, distinct roots, where $\overline{W} = X_0X_0'\mu^2 + \sum_{i=1}^{k+1} X_iX_i'\sigma_i^2$

$= J\mu^2 + V$, where V is the variance matrix of y in the corresponding completely random case and e) $P_iX_j(i\neq 0)\neq(j\neq k+1) = 0$ where P_i's were defined above then we have shown that the sufficient statistic $(X\hat{\gamma}_{LS}, s^2_{r+1},\ldots,s^2_{k+1})$ for the parameters $(X\gamma , \sigma^2_{r+1},\ldots,\sigma^2_{k+1})$ is complete. The counterpart of 2 above, namely b.l.u. estimators for estimable functions of regression parameters and b.q.u. estimators for variance components for the mixed model has been given. Analogous results under slightly more extended restrictions for a mixed model with $E(\beta_i\beta_i') = (a_i \backslash b_i)$ (i=r+1,...,k) have been presented.

The class of Model (VI.1) or (VI.2) situations with $E(\beta_i\beta_i') = I\ \sigma_i^2$ and for which for at least some i, j $(i\neq j)$, $X_iX_i'X_jX_j' = X_jX_j'X_iX_i'$ or the number of roots of W (or $\overline{W}$) exceeds k+2 we designate class S-P. In class S-P, for all examples thus far exhibited, even if normality of β_i's is assumed, the minimal sufficient set of statistics is not complete. There is no known general procedure for obtaining the M.V. estimator from a minimal set that is not complete. It is also not known whether or not UMV estimators exist when completeness does not hold. In view of these difficulties we find it both reasonable and practicable at present to discriminate among estimators on the basis of their computed variances. Techniques for finding variances of estimators of variance components are in our view inadequate; Chapters IV and V of this thesis attempt to increase our ability to calculate variances of variance component estimators obtained by the least squares method.

In Chapter IV we present a transformation method of obtaining estimators that is appropriate in "designed unbalanced" cases like the b.i.b. with random blocks and random treatments for example. In effect this "method" gives a single degree of freedom breakdown of the total S.S., and therefore gives us incidentally information on whether the components that usually go to make up a "line" of a conventional A.o.V. table when we use least squares, are homogeneous. This is desirable if we are to form weighted estimators since usually we weight inversely as the variances. Furthermore having available the matrix of the transformation allows us to obtain variances of variance component estimators quite simply.

In Chapter V we attack the general problem of the variance of a quadratic form, and the covariance between two forms that arise in mixed and random models. We find considerable simplifications in the case when we use a least squares method of estimation that is also known as Henderson's Method 3, and simplification further is possible if we assume normality of random effects. We have applied these results to obtain variance formulae for S.S.'s which have been suggested for finding estimators of variance components in random models with added concomitants.

VII. BIBLIOGRAPHY

Aitken, A. C. 1934. On least squares and linear combinations of observations. Royal Society of Edinburgh Proceedings 55: 42-48.

Bartlett, M. S. 1937. Some examples of statistical methods of research in agriculture and applied biology. Royal Statistical Society Supplement to the Journal 4: 137-183.

Bennett, C. A. and Franklin, N. L. 1954. Statistical analysis in chemistry and the chemical industry. New York, N. Y., John Wiley and Sons, Inc.

Bhattacharyya, A. 1945. A note on the distribution of chi-squares. Sankhya 7: 27-28.

Blackwell, D. 1947. Conditional expectation and unbiased sequential estimation. Annals of Mathematical Statistics 18: 105-110.

Bross, I. 1950. Fiducial intervals for variance components. Biometrics 6: 136-144.

Bush, N. and Anderson, R. L. 1963. A comparison of three different procedures for estimating variance components. Technometrics 5: 421-440.

Cox, D. R. 1956. A note on weighted randomization. Annals of Mathematical Statistics 27: 1144-1151.

Cramer, H. 1946. Mathematical methods of statistics. Princeton, N. J., Princeton University Press.

Crump, S. L. 1947. The estimation of components of variance in multiple classifications. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Crump, S. L. 1951. The present status of variance component analysis. Biometrics 7: 1-16.

Daniels, H. E. 1939. The estimation of components of variance. Royal Statistical Society Supplement to the Journal 6: 186-197.

David, F. N. and Johnston, N. L. 1951. A method of investigating the effect of non-normality and heterogeneity of variances on tests of the general linear hypothesis. Annals of Mathematical Statistics 22: 382-392.

David, F. N. and Johnston, N. L. 1952. Extension of a method of investigating the properties of analysis of variance tests to the case of random and mixed models. Annals of Mathematical Statistics 23: 594-601.

Dayhoff, E. E. 1964. Generalized polykays and application to obtaining variances and covariances of components of variation. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Dayhoff, E. E. 1964. On the equivalence of polykays of the second degree and Σ's. Annals of Mathematical Statistics 35: 1663-1672.

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. Biometrics 3: 1-21.

Fisher, R. A. 1925. Theory of statistical estimation. Cambridge Philosophical Society Proceedings 22: 700-725.

Fisher, R. A. 1930. Moments and product moments of sampling distributions. London Mathematical Society Proceedings 30: 199-238.

Gates, C. E. and Shiue, C. 1962. The analysis of variance of the S-stage hierarchal classification. Biometrics 18: 529-536.

Gautschi, W. 1959. Some remarks on Herbach's paper, 'Optimum nature of the F-test for Model II in the balanced case'. Annals of Mathematical Statistics 30: 960-963.

Golub, G. H. 1963. Comparison of the variance of minimum variance and weighted least squares regression coefficients. Annals of Mathematical Statistics 34: 984-991.

Gower, J. C. 1962. Variance component estimation for unbalanced hierarchical classifications. Biometrics 18: 537-542.

Graybill, F. A. 1954. On quadratic estimates of variance components. Annals of Mathematical Statistics 25: 367-372.

Graybill, F. A. and Hultquist, R. A. 1961. Theorems concerning Eisenhart's Model II. Annals of Mathematical Statistics 32: 261-269.

Graybill, F. A. and Weeks, D. L. 1959. Combining interblock and intra-block information in balanced incomplete blocks. Annals of Mathematical Statistics 30: 799-805.

Graybill, F. A. and Wortham, A. W. 1956. A note on uniformly best unbiased estimators for variance components. American Statistical Association Journal 51: 266-268.

Hammersley, J. M. 1949. The unbiased estimate and standard error of the intraclass variance. Metron 15: 189-205.

Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics 9: 226-252.

Herbach, L. H. 1957. Optimum properties of analysis of variance tests based on Model II and some generalizations of Model II. New York University Engineering Statistics Laboratory Scientific Paper No. 6.

Hooke, R. 1956a. Symmetric functions of a two-way array. Annals of Mathematical Statistics 27: 55-79.

Hooke, R. 1956b. Some applications of bi-polykays to the estimation of variance components and their moments. Annals of Mathematical Statistics 27: 80-98.

Hsu, P. L. 1938. On the best unbiased quadratic estimate of the variance. London University Statistical Research Memoirs 2: 91-104.

Hultquist, R. A. and Graybill, F. A. 1965. Minimal sufficient statistics for the two-way classification mixed model design. American Statistical Association Journal 60: 182-192.

Huzurbazar, V. S. 1963. Notes on the Theory of Estimation. Unpublished lectures. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.

Imhof, J. P. 1960. A mixed model for the complete 3-way layout with two random-effects factors. Annals of Mathematical Statistics 31: 906-928.

Kapadia, C. H. 1962. Minimal sufficient statistics for the partially balanced incomplete block (PBIB) design with two associate classes under an Eisenhart Model II. Institute of Statistical Mathematics Annals 14: 63-71.

Kapadia, C. H. and Weeks, D. L. 1963. Variance components in two-way classification models with interaction. Biometrika 50: 327-333.

Kempthorne, O. 1952. The design and analysis of experiments. New York, N. Y., John Wiley and Sons, Inc.

Kendall, M. G. and Stuart, A. 1961. The advanced theory of statistics. Vol. 2. London, England, Charles Griffin and Company, Limited.

Koopman, B. O. 1936. On distributions admitting sufficient statistics. American Mathematical Society Transactions 39: 399-409.

Lancaster, H. O. 1954. Traces and cumulants of quadratic forms in normal variables. Royal Statistical Society Journal Series B 16: 247-254.

Lehmann, E. L. 1950. Notes on the theory of estimation: monograph. Berkeley 4, California, University of California.

Lehmann, E. L. and Scheffè, H. 1950. Completeness, similar regions and unbiased estimation. 1. Sankhya 10: 305-340.

Lehmann, E. L. and Scheffè, H. 1955. Completeness, similar regions and unbiased estimation. 11. Sankhya 15: 219-236.

Magness, T. A. and McGuire, J. B. 1962. Comparison of least squares and minimum variance estimates of regression parameters. Annals of Mathematical Statistics 33: 462-470.

Manamunulu, D. M. 1963. Sampling variances of the estimates of variance components in the unbalanced 3-way nested classification. Annals of Mathematical Statistics 34: 521-527.

Matern, B. 1949. Independence of non-negative quadratic forms in normally correlated variables. Annals of Mathematical Statistics 20: 119-120.

Mirsky, L. 1961. An Introduction to Linear Algebra. Oxford, England, Clarendon Press.

Pearson, K. 1933. On the application of the double Bessel function $Kt_1Kt_2(x)$ to statistical problems. Biometrika 25: 158-217.

Plackett, R. L. 1960. Regression analysis. Oxford, England, Clarendon Press.

Rao, C. R. 1945. Information and accuracy attainable in the estimation of a statistical parameter. Calcutta Mathematical Society Bulletin 37: 81-91.

Rao, C. R. 1947a. General methods of analysis for incomplete block designs. American Statistical Association Journal 42: 541-561.

Rao, C. R. 1947b. Minimum variance and the estimation of several parameters. Cambridge Philosophical Society Proceedings 43: 280-283.

Rao, C. R. 1949. Sufficient statistics and minimum variance estimates. Cambridge Philosophical Society Proceedings 45: 213-218.

Rao, C. R. 1952. Some theorems on minimum variance estimation. Sankhya 12: 27-42.

Rao, C. R. 1962a. Apparent anomalies and irregularities in maximum likelihood estimation. Sankhya Series A 24: 73-101.

Rao, C. R. 1962b. Note on a generalized inverse of a matrix with applications to problems in Mathematical Statistics. Royal Statistical Society Journal Series B 24: 152-158.

Robertson, A. 1962. Weighing in the estimation of variance components in the unbalanced single classification. Biometrics 18: 413-417.

Roy, S. N. 1957. Some aspects of multivariate analysis. New York, N. Y., John Wiley and Sons, Inc.

Roy, S. N. and Cobb, W. 1960. Mixed model variance analysis with normal error and possibly non-normal other random effects 1. The univariate case. Annals of Mathematical Statistics 31: 939-957.

Roy, S. N. and Gnanadisikan, R. 1959. Some contributions to ANOVA in one or more dimensions. 1. Annals of Mathematical Statistics 30: 304-317.

Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. Biometrics 2: 110-114.

Scheffè, H. 1956. A mixed model for the analysis of variance. Annals of Mathematical Statistics 27: 23-36.

Scheffè, H. 1959. The analysis of variance. New York, N. Y., John Wiley and Sons, Inc.

Schopf, A. H. 1960. On the Kantorovich inequality. Numerische Mathematik 2: 344-346.

Searle, S. R. 1956. Matrix methods in components of variance and covariance analysis. Annals of Mathematical Statistics 27: 737-748.

Searle, S. R. 1958. Sampling variances of estimates of variance components. Annals of Mathematical Statistics 29: 169-178.

Searle, S. R. 1961. Variance components in the unbalanced 2-way nested design. Annals of Mathematical Statistics 32: 1161-1166.

Thompson, W. A., Jr. 1962. The problem of negative estimates of variance components. Annals of Mathematical Statistics 33: 273-289.

Thompson, W. A., Jr. and Moore, J. R. 1963. Non-negative estimates of variance components. Technometrics 5: 441-449.

Tukey, J. W. 1949. Dyadicanova: an analysis of variance for vectors. Human Biology 21: 65-110.

Tukey, J. W. 1950. Finite sampling simplified. Princeton University Statistical Research Group Memorandum report 45.

Tukey, J. W. 1956. Variances of variance components. 1. Balanced designs. Annals of Mathematical Statistics 27: 722-736.

Tukey, J. W. 1957. Variances of variance components. 2. The unbalanced single classification. Annals of Mathematical Statistics 28: 43-56.

Tukey, J. W. 1962. The future of data analysis. Annals of Mathematical Statistics 33: 1-67.

Wald, A. 1940. A note on analysis of variance with unequal class frequencies. Annals of Mathematical Statistics 11: 96-100.

Wald, A. 1941. On the analysis of variance in multiple classification with unequal class frequencies. Annals of Mathematical Statistics 12: 346-349.

Wald, A. 1947. A note on regression analysis. Annals of Mathematical Statistics 18: 586-588.

Watson, G. S. 1955. Serial correlation in regression analysis. 1. Biometrika 42: 327-431.

Weeks, D. L. and Graybill, F. A. 1961. Minimal sufficient statistics for balanced incomplete block design under an Eisenhart Model II. Sankhya Series A 23: 261-268.

Weeks, D. L. and Graybill, F. A. 1962. A minimal sufficient statistic for a general class of designs. Sankhya Series A 24: 339-354.

White, R. F. 1963. Randomization analysis of the general experiment. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Wilk, M. B. and Kempthorne, O. 1955. Fixed, mixed and random models. American Statistical Association Journal 50: 1144-1167.

Wilk, M. B. and Kempthorne, O. 1956. Some aspects of the analysis of factorial experiments in a completely randomized design. Annals of Mathematical Statistics 27: 950-985.

Yates, F. 1940. The recovery of interblock information in balanced incomplete designs. Annals of Eugenics 10: 317-325.

Zyskind, G. 1958. Error structures in experimental designs. Unpublished Ph.D. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa.

Zyskind, G. 1962. On conditions for equality of best and simple linear least squares estimators. Paper presented at Statistical Meetings, Minneapolis, Minnesota. Abstracts booklet available from American Statistical Association, 810 18th Street, N.W., Washington 6, D. C.

Zyskind, G., Kempthorne, O., White, R. F., Dayhoff, E. E. and T. E. Doerfler. 1964. Research on analysis of variance and related topics. Wright-Patterson Air Force Base Aeronautical Research Laboratory Monograph 64 - 193.

VIII. ACKNOWLEDGMENTS

All investigations on which this thesis is based were supported by the United States Air Force under Contract No. AF 33-615-1737 monitored by Wright Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio.

The painstaking attention of Dr. George Zyskind and the advice of Professor Oscar Kempthorne in the preparation of the final draft of this thesis, did much to improve my version, and are gratefully acknowledged.

5

61